The Myth of Interactive Audio:

Game Sound Dichotomies and Implementation Strategies

Andrew Smith

MA research report submitted in partial completion for masters of arts by coursework and research report

University of Witwatersrand

Supervisor - Hanli Geyser

11 February 2014

Amended 25 May 2014

The Myth of Interactive Audio: Game Sound Dichotomies and Implementation Strategies

Abstract

The audio development cycle for popular, mainstream digital games is currently a long-winded and highly convoluted undertaking, requiring several teams of full time audio specialists, composers and audio programmers for each release. This research report surveys the numerous advantages of sound in games, but also highlights the fundamental intractability of implementing credible soundscapes in interactive contexts. It examines why game audio practitioners are required to achieve far more with much less than their brethren working in linear media forms such as film and TV. This paper argues that recorded sound and interactive virtual worlds are uncomfortable bedfellows at best and, in some cases, can be thought of as mutually incompatible. As a result, many game audio strategies are little more than cunning attempts at bludgeoning temporally resistant audio files into interactive contexts. The report also explores a number of innovative sound design and audio implementation solutions for use in the digital games of the not too distant future.

Table of contents

Introduction

After every successful battle in *Bioshock Infinite* (Irrational), an ascending flurry of short string notes punctures the preceding battle music, thereby informing the player of a new non-menacing game state. The music subsequently recedes into a blissful oasis of sonic calm and your companion, Elizabeth, verbally communicates the current game status or upcoming tasks, while simultaneously furthering the narrative. In the same game, an effervescent popping sound confirms health replenishment and the sound of a cash register signifies the successful acquisition of the in-game currency. In all cases, the sound design, dialogue and music propel the narrative thrust of the game, ensure that the player is informed and up to date on game states, and engulf them in a believable, affective virtual world. The audio components of digital games are thus a highly efficient and emotive means for players to process, make meaning of and empathise with the game world.

The process of creating game soundtracks has dramatically increased in complexity over the last two decades as “the tools and talents of those who create them are indeed light years ahead of where they started” (Marks and Novak 19). In order to create a convincing virtual reality, game sound practitioners go to great lengths to manipulate the sonic properties of the games audio assets interactively. Moreover, interactive sound in digital games requires a host of unique skill sets and thought processes to help game audio practitioners overcome its multifarious idiosyncrasies.

In the interest of brevity, a number of restrictions are by necessity placed on this work. To that end, the principles of sound engineering and recording techniques are either avoided or treated superficially. The study is largely non-technical and therefore a rudimentary understanding of sound and recording technology is expected as a pre-requisite. Game construction relies on variances in space and time and, in this paper, these spatial and temporal properties are addressed in terms of simple Cartesian coordinates with finite dimensions. In other words, the metaphysical implications of space and time largely take a back seat in favour of structural analyses, as these aspects are more relevant to the complexity of practical game sound development. Although music is highly engaging and forms a key component of the powerfully affective and immersive properties of game audio, music composition techniques are trivialized herein. Instead, this work reviews the hypnotic and emotive advantages of music in addition to its structural and temporal disadvantages. Digital games often include cinematics and cut-scenes. These pre-scripted and animated videos are constructed in a linear, sequential and predictable fashion similar to that of film and TV practice. Well implemented cinematics can provide a powerful opening for a game, set the scene for what is about to follow, and provide reasons why you should care. Cut-scenes refer to the frequent dramatic intermissions between the interactive action, which further the narrative, reward players for completing a level, and introduce the next one. Despite their widespread use in digital games, however, cinematics are temporally fixed, and, “since these mini-movies are non-interactive” (Marks and Novak 72), they are not the focus of this paper.

Digital games with a science fiction theme were chosen as a focussing and contextualizing mechanism for this research report for a number of reasons. Firstly, science fiction was a primary instigator of the advancement of sound design techniques, audio technologies and “sound signification in film (narratively, thematically and aesthetically)” (Whittington 4) and will no doubt continue to push these boundaries in all forms of audio-visual entertainment in the future. Secondly, because the fictional worlds devised for science fiction games usually have no real world counterpart, a considerable amount of sound design expertise is required to believably bring them to life. William Whittington notes that “science fiction poses particular problems in terms of sound and image relations in that many of the environments, technologies and creatures do not exist” (106). Thirdly, the science fiction theme in digital games is extremely well represented by modern releases, with a profusion of culturally significant titles being published over the last decade. Fourthly, I am intrigued by the manner in which science fiction games are often able to challenge key contemporary concerns such as utopia versus dystopia, disenfranchisement and man versus machine whilst simultaneously acting as a lens for illuminating views of a possible future for society (Whittington 5). In many cases science fiction allows us to unmask modern socio-political problems and take them to their logical but dystopic extreme. Thematically ambitious digital games such as those found in the *Deus Ex* and *Mass Effect* series feature complex narratives which provide ambiguous takes on ideology, morality and choice while philosophical issues such as wealth disparity, trans-humanism, media control and excessive corporate independence are all expertly interwoven with the action. Lastly, science fiction soundtracks most commonly exhibit the phenomenon that I refer to as ‘noisic’ – a functional, intimate and ludicly useful interaction between sound effects (noise) and music. Science fiction works therefore provide an ideal vehicle for the discourse surrounding the complexities of game audio systems and often epitomize the multifaceted nature of game sound design.

Chapter 1

Game Audio: Film Sound, and then some…

1.1 A Game Audio Lexicon

Despite a number of similarities between sound for films and sound for digital games, the film and TV sound lexicon is often unable to fully service the needs of game audio studies and many definitions have yet to be agreed upon. Karen Collins admits that “as a discipline, the study of games is still in its infancy, struggling through disagreements in terminology and theoretical approach” and that “the fact that game studies is such a recent endeavour means that much of the needed empirical evidence has not yet been gathered or researched” (Game Sound 2). Despite such lexical confusion, this section describes a number of useful audio-visual, sound design and game audio terms for the purpose of this paper.

The physical dimensions of sound, which are equally applicable to all audio visual-media, can be described in terms of their frequency, amplitude, time, timbre and the space that they occupy (Moylan 5, 6). ‘*Frequency*’ is commonly referred to as pitch in musical contexts and, as a result, is one of the most recognized of all sound terms. Each note of a melody can be distinguished from the others because it operates at a different frequency to its collaborators. ‘*Amplitude*’ is colloquially known as volume and can be thought of as the perceived loudness or quietness of sounds. ‘*Time*’, a fundamental property of sound and an inescapable dilemma for game sound practitioners, is discussed in greater detail in Section 1.3. ‘*Space*’ or ‘*spatiality*’ is the “impression of the physical location of a sound source in an (…acoustic…) environment, together with the modifications the environment itself makes on the sound” (23). The aesthetic implications of technologies which simulate the spatialization of sound in this manner are discussed more fully in Chapter 3.

The term ‘*Timbre*’ requires additional consideration, as it is a critical component of game sound design, and thus referred to throughout this work. Timbre is also known as the ‘spectral content’, ‘tone quality’ or ‘tone colour’ (Bordwell and Thompson 185) of a sound, and is “the composite of all of (…its…) frequency components” (Moylan 6). Howard and Angus offer a number of “timbre to sound quality descriptions” such as “mellow, rich, dull, bright, dark, strident, harsh and shrill” (232). If we exclude the space and time dimensions for one moment, Floyd E. Toole summarises it thus: “In the perception of sound qualities, timbre is what we have left after we have accounted for pitch and loudness” (141). To illustrate the importance of timbre in music, consider an acoustic guitar string which, when plucked at a certain pitch, does not emit one single frequency but hundreds of them. The frequency that we perceive as the pitch or musical note of the instrument is typically the loudest and the lowest frequency present and is known as the ‘*fundamental*’ frequency. The additional frequencies that are above the fundamental, but specifically in tune with it, are known as ‘*harmonics*’. Generally, musical instruments, especially those with a readily identifiable melodic pitch, contain mostly harmonic frequencies with musically related intervals between them. Harmonic frequencies also have a simple integer relationship to the fundamental which leads to them being exact mathematical multiples of it (Izhaki 211) and (Moylan 7).

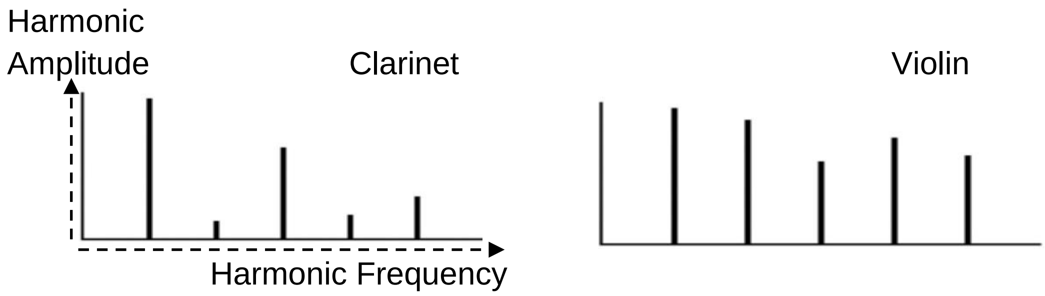

Figure 1. Differences between the harmonic content of a clarinet and a violin

As additional examples of musical, harmonic or tonal sounds, Figure 1 provides an approximate outline of the first five harmonics of the clarinet and violin respectively. The graphs are displayed with frequency components on the x or horizontal axis and the amplitude of each harmonic on the y or vertical axis. In both cases, the frequency of the first harmonic or fundamental is identical, and both contain primarily harmonically related frequencies. The differences in the relative amplitudes of the harmonics however, enable us to audibly distinguish one instrument from the other (Toole 141). Consequently, the fundamental frequency most strongly informs us of which musical note or pitch is playing, while the remaining harmonics validate which instrument is playing.

In addition to musical tones, the concepts of timbre and spectral content apply equally to all perceived sounds, including sound effects and voice. The term 'in-harmonics' is often used to refer to the myriad frequencies that are out of tune with, or "not proportionally related to, the fundamental" (Moylan 7,8). Sounds which contain predominantly in-harmonics are perceived as atonal or pitch-less. For example, it is not possible to whistle or hum a gunshot melody, as this sound consists principally of additional frequencies that are not musically related to each other. Such sounds tend to be “un-pitched, noisy, buzzy or clangorous” (Russ 224) rather than tonal, and familiar real world examples include wind, rain, thunder and explosions. Similarly, many musical sounds such as cymbals, snares and most percussion instruments consist primarily of in-harmonics rather than harmonics (Izhaki 211). When sung, words contain mostly harmonic frequencies, but when whispered, are principally in-harmonics or noise. Speech is a little more ambiguous in this regard as it contains both harmonics and in-harmonics, or, to put it colloquially, tone and noise respectively. I refer to the blurring of the distinctions between sound effects, voice and music in games as ‘noisic’, and this timbral and diegetic superimposition is discussed in more detail in Section 2.5.

Another sound design term which requires further clarification is ‘*sampling*’, which, at its simplest, “means making a digital recording of a relatively short sound” (Roads 117). These real world recordings commonly include musical notes, melodies, or full orchestrations. Moreover, single words, phrases and sentences are all easily sampled, as are sound effects and any sequential combinations thereof. The terms sample, wave file, audio file and digitized recording are often thought of as virtually synonymous. Nonetheless, samples in particular, are audio files with a comparatively brief duration and are usually triggered in real time as an audio event. A ‘*sampler*’, in music and sound design nomenclature, for instance, refers to “a piece of electronic musical equipment that records a sound, stores it and then replays it on demand” (Russ 316). This implies that the sample is triggered from a keyboard but, in game audio contexts, samples are the discrete digital audio files which are activated via game engine code hooks. Sampled sound is currently a ubiquitous practice for the creation ingestion, implementation and playback of game audio assets (Farnell, Designing Sound 318). Note however, that even electronic music scores which make use of all synthesised instruments, are still mixed, delivered and implemented in console and PC games as samples. In other words the synthesizing of these electronic sounds takes place in the composer's studio, not at runtime from within the game engine.

'Linear media' and '*non-linear media*' are an important pair of terms that are widely used to discriminate between chronologically inclined media and the indeterminacy of interactive media forms. Film, video, TV and music are commonly referred to as linear media since they enjoy fixed durations with sequential and permanent temporal structures prescribed by their creators. The developers of assets for interactive or non-linear media such as digital games, however, are afforded no such luxury due to the inherent unpredictability of the medium. Although sometimes used interchangeably, the terms '*interactive*', '*adaptive*' and '*non-linear*' audio usually describe the myriad attempts by game audio developers to "adapt (…the sampled audio assets…) to what the player is experiencing in a game" (Marks and Novak 22). These terms are widely used in game audio discourse but many game sound theorists have yet to reach complete consensus as to their correct use. Aaron Marks, for instance, refers to audio which 'appears' able to "follow the unpredictable and ever changing onscreen action" as '*adaptive audio*' and claims that it is "quickly establishing itself as a new standard" (6). Collins, on the other hand, prefers '*dynamic audio'* as a term to describe any sound that has the ability to change. She further suggests that dynamic audio can be subdivided into 'a*daptive audio',* which reacts to gameplay states, and '*interactive audio',* which reacts to the player directly via the controller. (Game Sound 184). This paper argues however, that these terms

are often used spuriously because the ideal interactive audio model for games has, in many ways, not yet been fully realised.

David A. Clearwater defines the term ‘*genre’* as a “method of categorization used to better understand cultural artefacts” but warns that it is “used and deployed in a variety of different ways” (Clearwater). In digital games for instance, the term is used to categorise gameplay types, mechanics and systems, whereas genre in linear audio-visual media forms is more commonly a classification system which defines narrative themes. Gameplay, interactivity and challenge are thus prioritized in the discourse surrounding game genre while “thematic or iconographic genres, such as Science Fiction, Fantasy and Horror, are (…) seen as little more than window dressing” or as “narrative enablers” and “the predominance of interactive genres over thematic ones is ascertained” (Arsenault).

While the games chosen for this study all employ science fiction as a narrative theme, they include a diversity of gameplay genres. For example, the mechanic which best describes *Bioshock Infinite* (2013) is a First Person Shooter (FPS), *Deus Ex: Human Revolution* (2011) could be considered a conflation of Stealth Action and First Person Shooter, *Mass Effect 3* (2012) is a Role Playing Game (RPG) commingled with a Third Person Shooter and *Portal 2* (2011) is a Puzzle Platformer. *Dishonoured* (2012) can be thought of as a fusion of Stealth, Action and Adventure, while *Dead Space 3* (2013) and *Metro Last Light* (2013) are both Survival Horror Shooters. Note that the recent trend of games which combine multiple gameplay elements allows many of them to willingly bask in the generic crossfire between shooting, platforming, puzzling and role playing mechanics. The list of genres represented in this research is far from exhaustive and several have been excluded, either because they were insufficiently represented within the realm of science fiction themed digital games, or were unable to effectively demonstrate the game sound strategies discussed in this paper.

‘*Diegetic*’ and ‘*non-diegetic*’ sound are another important pair of audio-visual terms which have been widely used in theatre and film discourse for decades. Diegetic sound is “sound whose apparent source is in the space-time continuum of the scene onscreen, (…) sound that the film leads us to believe that the characters can hear” (Chion, Film, A Sound Art 474). Within the context of digital games, both in-game characters and the player perceive diegetic sound. Diegetic sounds include dialogue, synchronised sound effects and background environmental effects, all of which are discussed in more detail in Chapter 2. Non-diegetic sound is that which is external to the diegesis of the game world: the characters are not aware of its existence and it is only heard by the player. Examples include the music score, voice-over narration, and most interface sounds which provide aural feedback, for the player’s ears only, of an in-game action.

Collins notes that “the notion of diegesis, borrowed from film and drama studies, is perhaps ill suited to games” because dynamic and interactive audio, which are integral components of game sound design, “complicate the traditional diegetic – non-diegetic division of film sound” (Game Sound 125). Kristine Jørgensen concurs and suggests that two categories of diegesis are insufficient for game audio studies since the player, who is external to the virtual world; and the player’s avatar (the diegetic personification of the player within the game), are inextricably linked via the game controller (107). Non-diegetic music that informs the player of a potential threat, for example, also has a direct effect on the avatar within the game, which can be safely moved away from the cause of the pending danger if desired. Theoretically the avatar cannot hear the source of the sound but is still able to act upon it via the player/avatar interface. This interactive structure means that, in digital games, the boundaries of the diegesis are often extremely difficult to define. It is therefore evident that game sound studies requires new terms and phrases to describe the ludic interpenetration of fictional and extra-fictional audio worlds. Jørgensen offers the term "*trans-diegetic"* for "this emerging frame of communication*"* (Jørgensen) and the game audio lexicon will undoubtedly be supplemented by additional terms as it continues to evolve. The additional ramifications of trans-diegetic and 'noisic' effects in science fiction digital games are examined in more detail in Section 2.5.

There are five traditional avenues of sensation, sensory input channels or '*modalities*' – taste, smell, sight, sound and touch, but at present, only the later three have any relevance to digital games. Collins states that sound in games "is '*multi-modal'* – that is, it involves more than one modality" and that it "includes haptics and visuals" (Playing with Sound 22). In addition, the implications of multi-modal phenomena on game players are a relatively new field of study and "much empirical research and theoretical exploration needs to be undertaken to understand it" (144). Although there is little doubt that these three sensory channels mutually influence each other during gameplay, this paper will focus on how, and why, the aural modality is often called upon to compensate for the shortcomings of the other two.

1.2 Sound and Image

Historically, sound professionals in all audio-visual media disciplines have endured a situation where the sound design “has taken a back seat to image, in terms of resource allocation, allotment of production time and screen crediting” (Beauchamp 17). Such “*ocularcentrism”* (Collins, Playing with Sound 22) is due in part to imperfect sound capture and playback technologies which, in the past, were unable to convincingly deliver immersive experiences that could rival those of pictures. Because of this, Altman notes that “the image has been theorised earlier, longer, and more fully than sound” (171). This situation has however changed dramatically and, over the last few decades, “sound has not only rivalled the innovative imagery of contemporary Hollywood cinema, which is replete with visual spectacles and special effects, but in some ways sound has surpassed it in status and privilege because of the immergence of sound design” (Whittington 1). In addition, the thunderous sonic spectacles that have characterized film since the advent of high fidelity distribution and playback systems have affected us culturally, and our aural expectations have thus been raised. Consequently, gamers, who grew up experiencing Hollywood blockbusters in all their hyper-realistic and surround sound glory, often expect similar aural experiences when playing games.

Sounds is recognizably no longer the poor cousin of visuals and, with the technologies firmly in place, audio theorists and practitioners have explored more fully the practical and aesthetic implications of the sound/image relationship. This section presents sound as an often superior sensory channel to sight when used in audio visual contexts, and suggests that it provides a number of useful properties that our ocular sense is physiologically incapable of. These sonic abilities evolved from our aural survival faculties but have since been creatively exploited by audio-visual media producers and audiences for entertainment purposes.

From a practical perspective, the mere presence of sound when accompanying visuals enables another modality, which allows more, and more complex, information to be presented to the audience at the same time. Jørgensen notes that “since the utilization of sound frees the visual system from attending to a lot of data, such as written messages rolling over the screen; it becomes easier to process a lot of simultaneous information” (175). Marks and Novak, observing our psychological reaction to aural stimuli, observe that “we have been conditioned all our lives respond to sound – such as answering the telephone when it rings, going to the door when there is a knock, and running out of the classroom when the bell sounds” (101). Other examples of sounds which “elicit a specific response” include a “snarling dog, a siren, scream or loud bang” (101). Physiologically, David Sonnenschein suggests that “sound can affect our body temperature, blood circulation, pulse rate, breathing and sweating” (71). Another biological aspect of sound, and one that is often overlooked, is that it can never be turned off because, unlike our sight modality, we have no ‘ear-lids’. This ‘always on’ aspect of the auditive is a consequence of our inherent survival mechanisms but it helps ensure that we are continuously attentive to aural cues while inhabiting virtual worlds.

Sound possesses many emotive qualities which are invaluable in audio-visual contexts. Listening is a more “sensual” (Breitlauch) and “carnal” (Collins, Playing with Sound 22) modality than viewing because “the intuitive nature of sound, unlike the obvious, categorical presence of sight, allows our minds to create more internal images and relationships" (Sonnenschein 152). The "creative power of the auditive seems to be more directed to the subconscious" (Breitlauch) which implies that the aural modality involves the imagination to a far greater degree than the ocular. Looking at this in another context, Stevens and Raybould observe that, "when films are rated for age appropriateness, it is often the sound effects which are revised to make the action feel less violent" (276). Chion observes that "most falls, blows, and explosions (…seen …) on the screen, simulated (...visually…) to some extent or created from the impact of non-resistant materials, only take on consistency and materiality through sound". These powerful textural connections between the player and a believable soundtrack mean that sound "has the unique ability to extend the self into virtual space" (Collins, Playing with Sound 42). This is especially relevant when animated media is considered, as much of the perceived physicality and veracity of animated objects is substantially derived from their sound sources, which are commonly recordings of real-world objects.

Michel Chion uses the term "*added value*" to describe "a sensory, informational, semantic, narrative, structural, or expressive value that a sound heard in a scene leads us to project onto the image" (Film, A Sound Art 466). This means that "sound can actively shape how we interpret the image" (Bordwell and Thompson 181), but the opposite is also true – "the image in turn influences our perception of the sound" (Chion 466). Chion claims that we do not specifically listen to, or view, these works but rather that we 'audio-view' them (Film, A Sound Art 466) and Walter Murch phrases it thus: "we do not see and hear a film, we hear/see it" (Foreword XXI). This implies a multi-modal sensory experience which is highly effective in entertainment contexts.

Sound is a more spatially comprehensive modality than visuals and this envelopment process significantly informs the immersive aspects of audio that are contemplated in Section 2.3. Our field of vision is limited to roughly 180 degrees but sound is experienced in 360 degrees which means that our hearing is naturally able to exceed the spatial boundaries of our visual acuity. To fully appreciate a three dimensional, enveloping visual space in the real world, we need to physically turn our eyes and body, first towards the frontal visual plane, and then towards the rear. The ears however, can completely comprehend the nuances of the same three dimensional space without so much as a twitch of the head.

In audio-visual contexts, the limited physiological boundaries of the ocular modality are further restricted by the frame of the screen on which the content is viewed. Chion asserts that “what the image designates in the cinema is not content but container: the frame” and “remains perceivable and present for the spectator as the visible, rectangular delimited place of the projection” (Audio Vision: Sound on Screen 66). Digital games however, feature additional capabilities, which, unlike film, enable the player guide the avatar ‘into’ the screen and change the camera viewpoint. Despite this liberating perception of a three dimensional world, the frame of the screen remains the definitive visual boundary during gameplay. Sound, however, is not thus constrained, as “there is neither frame nor pre-existing container” (67). Multi-speaker surround sound systems are becoming ubiquitous in both cinematic and home theatre applications, but this phenomenon does not exist for visuals. Even if 360 degree ‘surround see’ displays were a technical possibility, the ocular modality would be unable to make use of them without constant and considerable physical movement.

Sound and music for moving images also fulfils a number of important technical functions such as hiding visual edits by placing continuous segments of audio across them. Daniel Chandler refers to this practice as a “sound bridge” where the audio “assists in making visual editing invisible” (170). Holman observes that “keeping sound constant across a picture cut implies that, although we may have changed point of view, we are nonetheless still in the same space” (162). In addition, sound and music can help support narrative continuity within a scene or level and act as a thematic arc throughout the audio-visual work. Other technical functions of the soundtrack include helping to define the geographic location and/or the historical era of the work, enhancing the mood and, especially in the case of music, “revealing unspoken thoughts and feelings of characters” (Davis 142, 143). Therefore, and even without the additional implications of interactivity discussed in this paper, sound is called upon to serve in a multitude of essential capacities in audio-visual media production.

1.3 Sound and Time

The above mentioned theories of sound and image relations illustrate that sounds are an extremely powerful tool, aesthetically, technically and affectively, for use in the construction of audio-visual entertainment works. The relations between sound and time, however, present countless challenges for interactive audio practitioners and a considerable number of these issues have yet to be fully resolved. Mechanically and structurally, all sounds are temporal phenomenon with “a sense of time that is linear and sequential” (Chion, Audio Vision: Sound on Screen 18). William Moylan observes that “sound only exists in time; sound can only be accurately evaluated as changes in states or values of the component parts, which occur over time” (162). As Justin Boyd describes, “sound is time tangible” and in ultimately “seems one of the few ways to experience time” (Marginallia), which implies an inextricable correlation between the two.

When describing the temporal nature of audio objects in sound design practice, the “overall shape of (…each discrete…) sound, plotted against time” (Russ 120), is defined in terms of an ‘*envelope*’. The ADSR envelope is one of the most commonly implemented envelope shapes in sound synthesis and is an acronym for Attack, Decay, Sustain and Release. These parameters are displayed using a simple two dimensional graph where time always occupies the x or horizontal axis, as shown in Figure 2. The y or vertical axis, on the other hand, is typically mapped variously to the amplitude, timbre and/or pitch of the sound to allow the temporal manipulation of these sonic parameters. The amplitude envelope displayed in Figure 2 shows the attack time as the time a sound takes to reach maximum amplitude at its onset and the decay time as the time it takes for this amplitude to decrease to the sustain level. It will be noted that the sustain level is the only non-temporal parameter of the four and is an unvarying value as long as the key of an electronic keyboard is depressed. Finally, the release time is the time it takes for the amplitude to fade to inaudibility after the key is released.

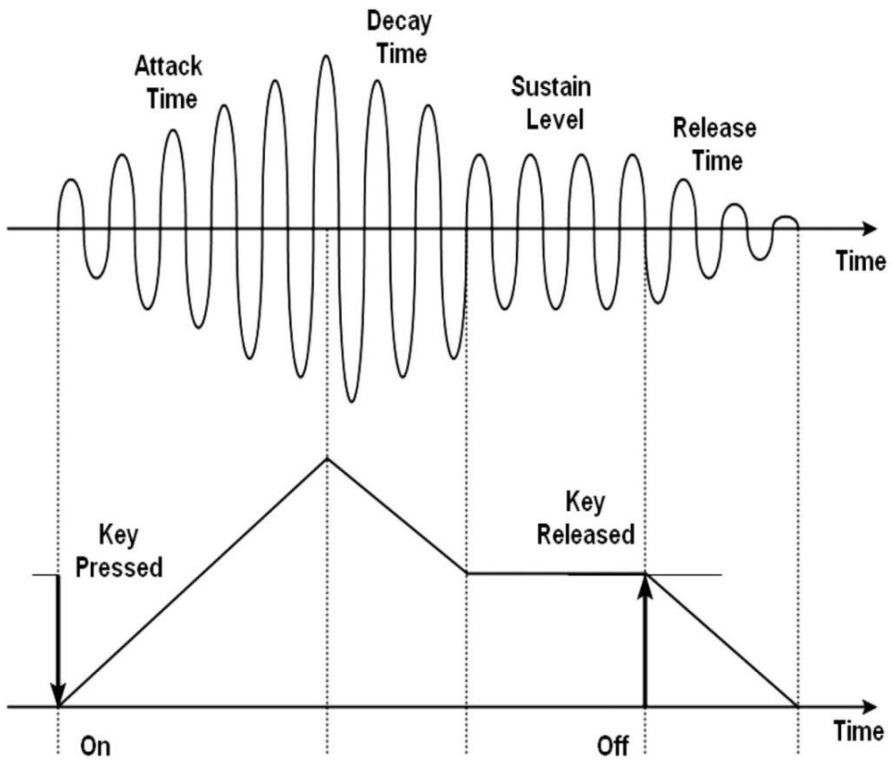

Figure 2. An envelope is a temporal attribute of all discrete sounds

The ADSR graph depicted in Figure 2 displays the amplitude variations of a single sound object, note or word over the duration of its existence. This particular ADSR envelope is artificially synthesized with onsets triggered via an electronic keyboard, but these concepts are readily adapted to describe the temporal vectors of any discrete sound in the natural world. To illustrate this, it is necessary to replace the term ‘key pressed’ in the graph with alternative real world phrases such as ‘note played’, ‘object struck’’ or ‘sound instantiated’.

Chion refers to each specific note of a piece of music as temporally “vectorized (…), with an irreversible beginning, middle and end” (19). A piano melody, for instance, are “composed of thousands of little indices of vectorized real time, since each note begins to die as soon as it is born” (20). Similarly, when an acoustic guitar string is plucked and held, it begins loudly and with a bright tone but becomes progressively softer in amplitude and timbre over time. This guitar note thus contains a fast attack time, a very slow decay time and no sustain. As another example, a flute note features a short bright burst or ‘chiff’ of noise at its onset which can be defined in terms of a fast attack and decay time. The instrument is then held at the sustain level for the duration of the note and, once the breath of the player has been removed from the instrument, the release time quickly decreases the amplitude to silence.

All real-world discrete sound effects and speech can also be considered in terms of temporal envelopes as they obey a similar sequential trajectory to individual music notes. An explosion, for instance, features a sudden onset of bright noise which thereafter decays in amplitude and timbre over a short period of time. These temporal aspects can thus be thought of as the amplitude and timbre envelope of the blast. On the other hand, the envelope of a creaky door varies little in amplitude or timbre over its duration, and its most notable temporal attribute is an often eerie variation in frequency or pitch. The sounds of a bullet ricochet and an engine revving are further examples of sounds which feature an envelope which manipulates pitch, albeit with radically different aural outcomes. Single words of dialogue are similarly encased within temporal envelopes as each word varies in amplitude, timbre and pitch over its duration. The syntactical, sequential and culturally agreed upon ordering of consonants and vowels in each word are an important factor in which they are made comprehensible. Consonants are commonly higher in amplitude and contain the brightest timbre while vowels contain fewer high frequencies and tend to softer in level. Depending on the position of the consonant within its makeup, each word thus evolves sonically over time.

Beyond discrete sound effects, musical notes and words, all of these aural objects are commonly arranged sequentially and in pre-defined patterns in order to make meaning. Speech for example relies on a forward temporal trajectory of consecutive syllable, word and phrase enunciations. With few exceptions, musical melody is also a linear media form which is metered according to pre-arranged and predictable chronological intervals. William L. Ashline states that “musical works are temporal and sonic objects that achieve their meaning through duration, just as objects in the visual arts achieve theirs through space” (Ashline). Tempo, an important component of most styles of music is measured in quantities of time known as beats per minute. In addition, a music score can be considered in terms of two dimensional Cartesian coordinates with an x axis for time and a y axis for pitch.

Edmund Husserl was periodically inclined to use musical analogies to illustrate his phenomenological theories of time consciousness and stated that “the tone in its duration is a temporal object” (The Phenomenology of Internal Time Consciousness 24). He further claimed that “each tone is a temporal extension itself” and that we are “conscious of the tone and the duration it fills (…as a…) continual flow” (25). Husserl’s phenomenological theories often make use of the phrase ‘a continuum of continua’ to describe time consciousness, but the term can also be used out of this context to describe the temporality of all sound. For example, all single notes, words and sound objects, no matter how short in duration, can be thought of as individual continua in that they always exist over a period of time. These continua are commonly sequenced into a continuum of sentences, melodies and soundscapes, or can be further extended in duration into full texts, musical compositions and audio-visual media productions. Another way of looking at this is to think of each discreet word, music note or sound effect as a micro-narrative which is assembled sequentially into the mini-narratives of a sentence, melody, or soundscape respectively and then further sequenced into complete narrative works. Although the term micro-narrative has also found use in other contexts, it appears perfectly suited to describe individual sound elements whose aural envelopes evolve over a period of time.

When considering sound in digital games, it is important to “make a distinction between events and processes in time, (…and…) the difference is that processes – unlike events – take time” (Kaae 76). The start, onset or instantiation of any sound can thus be thought of as an event, while the duration of that sound thereafter might best be considered a process. In game audio contexts, a triggered sample will simply play on for the duration that it was originally recorded because, “sampled sound is fixed in time” (Farnell, Designing Sound 318). This means that the beginning of a sample can be implemented interactively but “in the case of a single, one-shot sound (…such as a sample...), the game logic is unaware of the sound length” (Farnell, Behaviour, Structure and Causality in Procedural Audio 321).

To appreciate the profound implications of the temporal limitations of samples in digital games, consider a virtual sliding door which triggers a suitable recording of one second in duration. This pre-determined length of the sample could be considered appropriate for the approximate time it takes for the average person to open or close a typical sliding door, but the player might choose otherwise. If the speed of the door movement is interactively controllable, there is the possibility that the visual animation will no longer be synchronised with the sound of this action. If the player opens the door too quickly, for example, the sound will still be playing when the door is already (visually) fully open. Conversely, if the player chooses to dawdle and opens the animated door very slowly, the sample will have ended before the door is opened to its fullest extent.

To further illustrate the fundamental intractability of using temporally predisposed samples in game audio practice, consider the branching dialogue mechanic of *Mass Effect 3*. In this game, the player can choose whether to listen to the conversation of the NPC or omit it, as might be the case on a repeated play through. If the player decides to skip a segment of speech in the middle of a sentence, the game engine will simply stop the dialogue at that temporal juncture. Moreover, in the same game, there are many occasions when it is possible to trigger a segment of unrelated NPC dialogue while diegetic audio recordings of speech in the game world are still playing back. This too would not occur in real life as the character would presumably wait until the playback device had ceased before starting a new conversation.

Mass Effect 3 attempts a partial solution to the innate temporality of dialogue by implementing a brief volume fade at the end of a word, which crudely moderates the transition from speech to non-speech. Other games add a spatializing effect such as reverb to soften the end of a word and lengthen it slightly. While these techniques help to disguise the problem somewhat, they just as commonly draw attention to the problem. In addition, and no matter which strategy is used, the result remains unconvincing as, in the real world; humans do not fade their sentences mid-speech and, unquestionably, cannot be instantly teleported to a reverberant location at the end of a word. Such obvious sonic discrepancies in digital games: between sounds as they occur in the natural world, and those experienced in a virtual representation of it, have the potential to severely disrupt the player's sense of immersion.

Many games delay the result of an interactive trigger until the end of a sentence and then terminate the dialogue before the next one, which sounds more natural and helps ameliorate linguistic confusion. However, as a consequence of the often significant time discrepancy between the trigger and its result, this is no longer interactive audio because the audio does not interact with the player instantaneously. Similarly, and as elaborated upon in Section 3.3, sampled melodies are often implemented in games with a time delayed trigger which waits until the end of a musical phrase before changing it. In all cases, it is the pre-disposed temporality and inherent continuity of speech and music which present numerous irresolvable stumbling blocks for developers of interactive media.

Despite the numerous difficulties posed by the use of fixed duration samples in digital games, not all sounds suffer from temporal problems to the same degree. Many brief sound effects or discrete words are so short-lived that they have ceased to exist before their temporality becomes a hindrance to interactivity. In other words, fleeting percussive or ephemeral sounds, such as gunshots and transitory verbal shout-outs, expire so quickly that they can be considered events in game audio practice. Longer samples however, such as all sentences, soundscapes and melodies, must be thought of as processes that, although instantiated interactively from a game audio code hook, often need to play to their conclusion in order to provide the required meaning. A key game audio dilemma, therefore, is that while the onset of each sample can easily be triggered randomly and instantaneously, its sonic behaviour thereafter is considerably more difficult to control interactively.

In cinematic applications, Bordwell and Thomson describe one of the many fascinating aspects of the temporal nature of sound: “we can't stop the film and freeze an instant of sound, as we can (…to…) study a (…visual …) frame” (264). Chion confers and notes that “sound presupposes movement from the outset” and that it “bears a fundamentally different relationship to motion and stasis” (Audio Vision: Sound on Screen 9). In addition, when synchronized sounds and visuals do ‘*freeze*’, due to bandwidth constraints for example, it is the audio drop-out that is most often perceived as being the most objectionable. Under these conditions, the moving images usually pause for a few frames and briefly resemble a still photograph. This brief visual phenomenon is barely noticeable to an audience because an image nonetheless remains on screen. The accompanying sound track however is not so fortunate and, when frozen, immediately collapses into disturbing and brutal silence for a fraction of a second. In most cases this extremely tiny sliver of non-sound is considerably more conspicuous and disruptive than the visual freeze-frame of a similar duration. In other words, images can comfortably be motionless, as in the case of still photography and the above mentioned visual freeze frame, or just as easily of the moving variety. Audio, on the other hand, needs to be constantly moving in order to exist, which means that the ‘still sound’ is a physical impossibility.

From an audio-visual perspective, a temporally useful aspect of the aural modality is that “audiences process, assign meaning to and store audio information more rapidly than they do visual information" (Beauchamp 18) because "the ear analyses, processes and synthesizes faster than the eye" (Chion, Audio Vision: Sound on Screen 10). This means that a rapid sequence of short, discrete sounds can effectively be used to '*speed up*' the image and create the impression for an audience that far more activity is taking place in the virtual world than the visuals are able to convey.

It should be apparent that all recorded speech, music and sound effects cannot be anything other than mutually and indivisibly associated with time. As discussed in Section 1.1, most sound assets used in digital games are digitally recorded and temporally static samples. The implications of the use of fundamentally linear and sequential samples within the context of interactive, non-linear and randomly inclined digital games are profound. Despite this dichotomy, game audio practitioners have developed a multitude of ingenious workarounds and partial solutions to the innate temporality of samples, a number of which are discussed in Chapter 3.

1.4 Sounds in Games

Many of the advantages of sound in conjunction with visuals in linear audio-visual media apply equally to the use of sound in games. This "in some ways makes games a continuation of linear media, and in other ways distinguishes the forms" (Collins, Game Sound 5). The similarities between the two have allowed game sound designers to draw upon numerous practical and theoretical resources from the film world. Furthermore, recent advances in digital sound playback technologies have allowed modern digital game soundtracks to convey significantly more meaning and veracity than their forebears. As Marks observes, game sound “technology has advanced to the point where (…) surround sound, full orchestral scores, Hollywood sound effects for every tiny thing, celebrity character voices, narration and multi-player voice interaction” (377) are regularly included in modern mainstream titles. On the other hand “there are a number of distinctions in the way audio functions in games” (Collins, Game Sound 128). The most notable of these distinctions is the fact that a game typically requires a far greater number of sound assets than a film of a similar budget, but has far less storage space to put it in. This dichotomy is discussed in further detail in Chapter 3.

Despite the simultaneous technological improvements to both sound and visuals in digital games over the last few decades, the two disciplines have conceptually and structurally drifted apart. Farnell notes that “the present way of doing things (…sampled audio files…) forces a partition between the design of (…an…) objects visual and sonic properties” (Obiwannabe).To allow for interactivity, the visuals of a digital game have always been computer generated and remain that way, despite enormous advances in animation technologies. The reason for this is that it does not appear possible to coerce photographed or filmed moving images of real world actors into behaving in an interactive manner. As result, animated visuals remain an essential component of game design as they are coded rather than recorded, and can thus be more easily manipulated within non-linear environments. Recent developments in body and facial motion capture technology have produced impressive interactive animations of real world actors, but an actual actor is not shown on screen. In other words, no matter how ‘real’ the animated images appear to be, “the player is still looking at pixels, (…and…) the only ‘real’ sense fully experienced by the player is sound” (Marks 68). Beauchamp concurs: “animators create characters, not actors” and “it is the voice talent that brings the characters to life” (20).

The use of sounds in games has evolved along rather different lines to those of animated visuals. The earliest arcade and console games of the 1970s and 1980s featured little more than synthetic or electronically generated “bleeps and bloops” (Marks 3) and (Collins, From Pac-Man to Pop Music 12, 13). Synthesized sound is typically not as realistic as recorded sound and, in many respects, can be considered similar to animated images as both are artificially generated using computer technology. From the 1990s onwards however, digital audio technologies progressed to the extent that it became feasible to include sampled sounds in games. This was a critical turning point for game audio as samples are not synthetic, but real world recordings of tangible objects. Sampled sounds are “more realistic in terms of their auditory fidelity” (Collins, Playing with Sound 36) and thus provide for greater immersion and emotive involvement than synthetic renditions thereof. This perceived realism has ensured that samples have dominated game audio practice for the last two decades and “synthesized sound was relegated to the scrapheap” (Farnell 318).

Samples in games can be thought of as similar to filmed or recorded moving images as both comprise time which has been captured in a material environment. The designers of the animated images for digital games however, have continually prioritised interactivity over veracity, despite the gradual increments in the realism of such visuals. Conversely, game audio developers were forced to forego a certain amount of interactivity in order to achieve far greater verisimilitude. Herein resides a crucial game sound paradox. Sampled sound is physically and structurally temporalized and must follow a continuous trajectory in order to exist and make meaning. The animated and interactive aspects of digital games however, demand non-linear constructs, which produce instantaneous, random and '*on the fly*' changes. A simpler way of putting this is to say that visuals in games are animated, rather than recorded, while most sounds in games are recorded, rather than synthesized.

The linear nature of sound and music recordings make them ideal accompaniments for linear audio-visual media forms such as film and TV. As Stevens and Raybould claim, "One of the chief differences between writing sound for film (…) and games is that in games the sounds you want available to play instantaneously" (33). So while "the greatest entertainment advantage of video games over film is that the experience can change every time it is played" (Marks and Novak 22), the production cycle for interactive sound designers is significantly more complex. The practical outcome of these observations is that, while sampled sound is corporeally advantageous in digital games, it is also, in many cases, temporally disadvantageous. Therefore, if sound can be thought of as 'time-made-tangible', interactivity implies that time be made intangible to the suit the non-linear nature of the medium. This dichotomy informs a large number of game audio implementation strategies, which, in many cases, consist of convoluted attempts to force these temporally obstinate audio samples into randomly varying and interactive contexts.

Chapter 2

Noisic: Aesthetics of Game Audio

2.1 Sound Types and Sound Sources

As “in the traditional hierarchy of film sound” (Whittington 201), there are three fundamental families of sound elements in games – sound effects (noise), music and dialogue. Sanders Huiberts observes that these “distinct paths in the production process of game audio” focus on “the original source of the recorded material and less on how it functions within the game” (15). The functions of game audio however, are discussed later in this chapter. In certain low budget cases, a single talented sound designer may be required to fulfil all three roles and compose the music, create the sound effects, record the dialog and even perform some of it. However, mainstream commercial games have recently “experienced unsurpassed growth and diversity associated with career paths” (Marks and Novak xi). Consequently, many games require dedicated music, voice and sound design personal because “the roles played by sound professionals have become increasingly specialized” (Collins, Playing with Sound vii).

Game music is typically sourced from a composer working in a dedicated studio in order to stamp an original identity on the work. Alternatively, as “a low cost means of developing a score” (Beauchamp 50), there are myriad ‘production music’ or ‘mood music’ libraries available. However, the non-exclusive nature of music libraries, their potential familiarity and a number of rights holder issues ensure that they are rarely used in mainstream game audio practice. Similarly, game dialog needs to scripted, acted and recorded specifically for each work. Because of the inescapable narrative requirements of each individual game, every voice recording is unique and a 'dialog library', if such a thing were to exist, would be superfluous.

The sound effects or noise category can be further subdivided into backgrounds and synchronized effects. Backgrounds are also known as environmental effects or ambiences and are crucial in "bringing the game setting to life" (Marks and Novak 5). Ambiences are fairly lengthy and continuous recordings of noises such as traffic, urban soundscapes (with the obligatory barking of dogs) or the countryside. Such un-synchronized sound effects "are not associated with any specific object" (75) and can be thought of as "the equivalent of a sonic long shot (wind through trees or waves crashing on a beach)" (Whittington 137). The interior spaceship levels of '*Mass Effect 3*' for example; continuously bathe the player in an atonal sonic soup of indistinct verbal chatter, space air, rumble and other mechanical noises. In most cases, these sounds have no observable visual cause but they serve to provide added dimensions of significance, immersion and verisimilitude for the player. Synchronized or 'hard' effects tend to be shorter in duration and "are linked to (…the…) onscreen objects" of the game. Such transient or impactful sounds "typically bear a one-to-one relationship to picture" (Holman 164) and examples include gunshots, punches, footsteps and body falls.

Where budget and/or time allows, game sound practitioners will always try to record original sound effects either indoors or out in the field using a portable field recorder.

The indoor capturing process mostly takes place in an acoustically isolated recording studio, and this is “the most desirable method for capturing quality game sounds” (Marks and Novak 82) because of its controlled acoustic environment. The down side of the indoor approach, however, is that the recording of certain sound effects, such as breaking large amounts of glass or physically demolishing anything substantial, tends to be a messy and occasionally dangerous process. Field or remote recording implies an outdoor expedition where the challenges of extreme weather and equipment breakdown test the skill of the most experienced sound designer (83). Hard effects such as gunshots and explosions are frequently recorded at a firing range for safety reasons (81) while most natural backgrounds are by necessity recorded outdoors or ‘in the field’. The caveat here however, is that field recording “can be fraught with unpredictability” (82), and therefore requires extreme patience to prevent frequently occurring, but unwanted and extraneous noises.

Marks and Novak refer to Foley as “the art of creating general movement and object handling sounds that are in sync to on screen actions” (79). To be technically correct however, the factor that specifically delineates Foley from hard effects is that Foley is ‘performed’ to an edited and temporally locked version of the visuals (Holman 166). Professional Foley artists thus watch the finished cut of the action on a large screen, and ‘act’ the sound effects correspondingly on a dedicated Foley stage. Examples of effects which are performed or acted include “footfalls, body falls, hits, weapon handling, and other general sounds a character would make while moving through a scene” (Marks and Novak 79). When used in the context of linear, structured and pre-scripted game cinematics or cut-scenes, Foley can be acted in sync with the animated visuals in much the same way as it is in film practice. But because this temporally locked situation does not exist for the interactive components of games however, the term Foley in these contexts can be considered as something of a misnomer.

In many cases, neither time nor budget permits the sound designer to capture an original recording for every instance of sound that is used in the game. When such a dilemma arises, game audio developers often turn to sound effects libraries – vast repositories of high quality, “pre-recorded, easy to use, time-saving wonders” (Marks 49). Sound effects libraries are a relatively cost effective option for the sourcing of game audio assets and are available either online or on tangible media. For science fiction games, however, the choices remain somewhat limited, and those sound effects that are available are not always perfectly suited to the production at hand. Compounding this problem is the fact that “everyone has access to the same libraries and it can be tough to maintain any kind of originality” (Marks and Novak 84). In addition, “hearing the same sound effect used on television, radio and in other games tends to be disruptive to the immersive experience” (84). In such cases, it is the talent of the game audio practitioner that is called upon to conceal their origin. A proficient sound designer would “never be satisfied with stock library sound effects” (Marks 312) and would typically combine a number of library sounds or sonically manipulate them using software based audio processing tools. If successful, this process can produce new and original content from well-worn samples.

2.2 Sound, Causality and Substitution

Film editor and film sound practitioner Walter Murch notes that, in the real world, sound and picture present themselves as inextricably connected because sounds are "stuck like a shadow to the objects that cause them" (Foreword XIX). Conversely, the captured moving images in audio-visual media are rarely the causal agent of the accompanying sounds heard on playback. Rick Altman refers to this as "sound film's fundamental lie: the implication that the sound is produced by the image when in fact it remains independent from it" (46). There are a number of unavoidable reasons for this practice. Even in the case of live film production, where the possibility certainly exists for the simultaneous recording of certain sounds with visuals, there are a number of technical reasons why this is often impractical. It has been shown that one advantage of sound is that it completely surrounds us in a 360 degree arc, as opposed to the much narrower spatial experience of the ocular modality and especially, of framed visual displays. This phenomenon is a disadvantage when recording sound, as although a camera can easily pan away from unwanted visuals, even the most directional microphones cannot selectively and completely isolate the individual sound elements of a real world soundscape. To make matters worse, recording sound for a live film is often an unpredictably noisy affair, as a result of the use of various types of off-screen machinery required on set. When recording outdoors, this situation is compounded by the often unwanted sounds of nature or indiscriminate transportation noises in urban areas. In addition, Tomlinson Holman notes that "it is always possible to add reverberation and background ambient sound later, (…) but it is extremely difficult to reduce excess reverberation in particular, and (…background…) noise in general, that accompanies original recordings" (46). As another example, Noël Burch observes that “the possibilities of recording a comprehensible conversation in a car are rather slight” (201) and consequently sound effects and dialog are nearly always recorded separately.

The technical necessity of “constructing reality (as opposed to observing it)” (Altman 47) in film sound has yielded a number of significant aesthetic advantages for sound designers in all audio-visual media disciplines. While this practice is more time consuming than simultaneous, on-set audio recording, the resulting “re-association of image and sound” (Murch XIX), has proven powerfully effective and affective. With all sound sources perfectly isolated from each other, and in the absence of extraneous and distracting background noises, soundtrack creators have been provided with liberating creative control over the outcome of their work. In addition, sound exhibits great “flexibility with respect to causal identification”, because “’the spectator is quite tolerant when a sound does not resemble what one would hear in a given real situation” (Chion, Film A Sound Art 488).

Through years of experimentation, audio-visual sound designers have discovered that the sounds of many objects that are not causally related to the visuals on screen often act as better sounding substitutes than a recording of the real thing. Murch for example, notes that “walking on cornstarch (…) records as a better footstep in snow than (…a recording of…) snow itself” (XIX). From an ethical and legal perspective, certain sounds cannot be anything other than substitutes. Examples include breaking a celery stick instead of a human bone, hitting a leather jacket as a sonic replacement for punching someone’s face and stabbing almost any fresh vegetable instead of a person. The survival horror game *Dead Space 3* exemplifies this technique, as a number of audio recordings of a water melon being trampled were used as stand-ins for the sound of violently mutilating a Necromorph's head. (http://www.cnet.com/profile/sonyinsider/). In a more causal relationship to the visuals than is the norm, this same sound recording "can convincingly serve as the sound effect for a crushed watermelon in a comedy" but, more dramatically, as a replacement "for a head being blown to smithereens in a war film" (Chion, Audio Vision: Sound on Screen 22).

Sound substitutes with unusual and, at first glance, illogical origins are always possible within audio-visual relationships, as "the audience is disposed to accept, within certain limits, these new juxtapositions as the truth" (Murch XIX). Sonnenschein notes that "we seem to be wired for imposing order on our perceptions, as the brain will pull out patterns from chaos, even if none objectively exist" (99). In other words, the ear tends to be very forgiving, and craves the attachment of any sound with a synchronous visual image, however highly unlikely the source material. Chion refers to this phenomenon as synchresis, and defines it as "the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time" (Audio Vision: Sound on Screen 63). Chion later summarized this definition as "synchresis consists in perceiving the concomitance of a discrete sound event and a discrete visual event as a single phenomenon" (Film, A Sound Art 493). This is a particularly useful property of the sound-image relationship that has found great favour in many audio-visual contexts.

Another common film sound practice which has been readily adopted by game audio developers is ‘hyper-realism’ – the process of deliberately exaggerating sounds for entertainment purposes. In other words, sound is used to magnify the emotive impact of the virtual world rather than merely representing it. Such sounds “convey the intended meanings of the visuals without necessarily being based on (…sonic…) reality” (Chion, Film A Sound Art 478). This “stretched reality can be said to be pushing the limits of the real” (Holman 175) but audiences readily accept “a simulacrum of the real” and thoroughly enjoy sounds that are perceived as “more real than real” (Collins, Game Sound 208). Chion refers to this phenomenon as ‘rendering’ – a situation where often wildly inaccurate sounds are believed to be faithful to the audio-visual environment by the audience (Film, A Sound Art 488). What are the reasons for this seemingly deaf acceptance by audiences? Crucially important to the practicability of unusual sound substitutes is that audiences do not recognise the recordings of many sounds, as they have not experienced them in real life. Examples include plane or train crashes, stabbing someone or shooting heavy weapons. In the science fiction genre of course, the majority of sounds lack any real world familiarity and must be unnatural by default. Another reason for our acceptance of sound substitutes is that, “the codes of theatre, television and cinema have created very strong conventions determined by a concern for the ‘rendering’ more than for literal truth” (Chion, Audio Vision: Sound on Screen 108). Through exposure, we have become completely familiarised by these conventions, “and they easily override our own experience and substitute for it, becoming our reference for reality itself” (108). The sound of real world punches or gunshots for example, are commonly perceived as aurally anaemic in comparison to their excessively hyped film or game audio counterparts.

In the hyper-real sonic universe of digital games, audio rendering is used extensively to create intense feelings of power, frailty or fear in the player. Successful audio rendering epitomizes the sound design of many action games, which often emit hyper-realistic, non-diegetic and brutally rendered crunch noises to indicate a successful enemy takedown. In other words “sound especially will be called upon to render the situation’s violence and suddenness” (Chion, Audio Vision: Sound on Screen 113). While these sound effects could initially be construed as merely informative of a success state within the game, they also substantially enhance the sense of power and immersion, despite the highly implausible nature of the sounds themselves. Such sonic hyperbole is common in other forms of modern audio-visual media – especially action movies – but when combined with interactivity in games, new meanings are generated that transcend those of the original sound files.
Where do synthetic sounds fit into this picture? It has been noted that sampled or recorded sounds provide the most verisimilitude and that these can be further stylized to sound ‘more real than real’ or hyper-realistic. From this perspective, we could consider synthetic sounds to be ‘less real than real’ or ‘hypo-realistic’, unless they are specifically required to represent or symbolize certain aspects of the 8 bit gaming era. To a large extent, this explains why synthesized sounds have been superseded by samples in modern game audio practice.

The creative possibilities offered to film sound designers of the 21st century have allowed them to extend audio-visual attachments beyond the mundane and create a deliberate “metaphoric distance between the images of a film and the accompanying sounds” (Chion XXII). This is common film sound practice and can “open up a perceptual vacuum into which the mind of the listener must inevitably rush" (Murch 247). Murch puts it another way: "If the audience members can be brought to a point where they will bridge with their own imagination such an extreme distance between picture and sound, they will be rewarded with a correspondingly greater dimensionality of experience" (Stretching Sound to Help the Mind See). Therefore, the sounds are not always used to directly represent the image but, by deliberately distorting the relationship between them, can be constructed to signify a multitude of additional meanings.

Anthropomorphism is "the idea that all objects have the potential to display human characteristics" (Beauchamp 21) and game sound design is replete with examples of these effects. In audio-visual media contexts, human screams are added to enemy spaceship sounds to make them more terrifying, moving metallic objects are made to 'moan' as if in agony and wind 'howls' around fixed structures (Whittington 143). Sonnenchein notes that, "we respond so primarily to a snake hiss or gorilla roar, that these sonic events can infuse an associated character (…or object….) with a sense of deadliness and ferociousness" (180). This means many animal noises can be successfully combined with virtually any other sound effect so that the result is perceived as more menacing. For example, the Star Wars 'Tie-Fighter' fly-by sound originated as a mutated recording of a real world elephant trumpet blast (Sonnenschein 193), but still elicits powerful emotions in audiences in the context of its science fiction narrative.

Jørgensen notes that "human perception tends to have more confidence in its visual capabilities than its auditory" (On transdiegetic sounds in computer games 107) and this phenomenon is particularly useful in audio-visual media contexts. Games in the survival horror genre, for example, feature “environments that are (…) designed to limit the visual perception of the gamer” (Roux-Girard 193). In other words the cause of certain sounds is deliberately hidden from view as a means to ratchet up the tension. As Randy Thom observes: “starving the eye will inevitably bring the ear, and therefore the imagination, more into play” (Thom). This practice will also, however, invoke considerable fear in the player, as they are forced “to utilize their ears to survive in the nightmarish worlds in which they play” (Roux-Girard 193). Dark, shadowy visual environments thus epitomize this genre, with sinister tunnels and unlit spaceship corridors being the norm. Survival horror shooter *Metro Last Light* makes great use of this terrifying effect, as the vision of the player is further obscured by their virtual gas mask, which quickly mists up through condensation and becomes badly cracked over time.

2.3 Sound and Immersion

In her PhD thesis, “‘*What are Those Grunts and Growls Over There?’ Computer Game Audio and Player Action*”, Kristine Jørgensen sought to reveal the functionalities of sound in computer games, with focus on the relationship between sounds and events in the game (Jørgensen). In Jørgensen’s study, a number of participants played games with and without sound and she subsequently concluded that the lack of audio feedback and immersion negatively impacted on their ability to function within the game world. In another empirical study, Sanders Huiberts found that “sound can induce or contribute to the feeling of presence in the game”, and “engage the player while masking the sound of the user environment” (57). Such affective feelings of strong involvement or deep engagement with the virtual world, in addition to emotional attachment, identification and bonding with the game characters, are all mentioned as important facilitators of immersion by players (Huiberts 39 - 41). Part of the gratification of gaming is thus summarized as the experience of “transportation into the game world, absorption in the activity and identification with the situation” (49). This is often referred to as the ‘willing suspension of disbelief’ by audiences as, although we are aware that the game is a virtual world, we choose to suspend that notion temporarily for entertainment purposes.

The immersive aspects of digital games aspire to empower the player with the means to suspend sensory input from reality and surrender instead to an imagined role in the fictional realm of the game. As Collins describes, “one of the key functions of sound in games is to immerse us in the virtual world through the sense of sonic envelopment” (Collins, Playing with Sound 54). Although these “hypothetical spaces are conceptually separate from our own world” (Jørgensen 106), it is the audio assets of digital games that are often able to successfully evoke convincing in-game embodiment. Stevens and Raybould claim that “immersion in a game is when a player loses sense of the outside world and is, for a moment, ‘lost in the flow’” (277). This “suppression of all surroundings (spatial, audio-visual and temporal perception)” (Nacke and Grimshaw 270) is essential as external sounds “will invade and detract from the immersive experience” (Marks and Novak 68). An example of sound design significantly aiding immersion is the ‘super jump’ in the sandbox game *Saints Row IV* (2013). When activated, a noisy arpeggio accompanies your skyward trajectory, providing exhilarating sensations of rapid ascension and ultimately, the believable perception of altitude.

It has been noted that fictive digital game worlds are compromised by virtue of the fact that, like all animated media, the visuals are artificially constructed computer objects and feature, at best, reasonably acceptable approximations of the real world. The human voice and recordings of other tangible objects are therefore vital in providing game objects with personality and authenticity through the use of sound. The dialogue recordings and expressive skill of the actors provide a critically important human reference for the player and stimulate much of the empathy that he/she feels for the game characters. Similarly the sound effects and ambiences used in a game are often recorded in real world factories, offices and firing ranges (to name just a few remote recording possibilities). These samples “give material solidity to a graphic being” (Chion, Audio Vision: Sound on Screen 118) and are therefore indispensable in game audio contexts.

Music is arguably the ultimate immersive audio element as it “is one of the most powerful tools to affect how players feel in games” (Stevens and Raybould 161). Well realized non-diegetic theme music has the potential to direct our emotions and it often achieves this “in a completely subconscious way” (Marks and Novak 130). Music is capable of “effectively hypnotizing the spectator” (Sonnenschein 104), which is somewhat ironic given that the music score is the most inauthentic aspect of the entire soundtrack. For example, when a couple kiss in the material world, romantic violin melodies do not start to play in the background. Audiences of audio-visual media not only accept this irrational logic, they allow the music to make the kiss somehow seem more tender. Similarly, in a real war, ominous up-tempo Taiko drums do not portend a hail of bullets, but their presence in the virtual world dramatizes the action and increases excitement levels. When considered rationally and out of audio-visual contexts, it is hard to believe that we would willingly permit "music to diminish our critical faculties and allow us to be more susceptible to suggestion" (Sonnenschein 105). As a means to greater entertainment value however, we always "relinquish control, (…because…) peak experiences are invited – moments of great insight, surrender, wisdom, or awareness of profound human relationships (105). Non-diegetic theme music is therefore purposefully scored to manipulate the emotions of the player, and for "setting the overall mood and tone" (Marks and Novak 130) of the game. Music can also be diegetic, as is the case when it emanates from in-game radios and TV's, or through game characters playing musical instruments or singing (Stevens and Raybould 163). This is "more representative of reality than music which appears to come from nowhere" (170) and serves to inform the player about the game characters, "defining the kind of people they are" (164). Both diegetic and non-diegetic music thus provide the player with powerfully emotive, immersive and informative connections to virtual environments and effectively provide for the suspension of disbelief.

The score typically covers a wide variety of in-game moods, from sedate and exploratory to action packed, but "there is always a strong bond that links all the music together into a cohesive whole" (Rona 41). Stylistically therefore, every game score "needs to have its own personality" or, as Rona puts it, "a '*door*', a single unique way to give it focus and identity" (3, 42). Stevens and Raybould observe that, while melodies, harmonies and rhythms are important to the score, the “instrumentation remains the most immediately evocative” (164). The instrument palette brings “significant cultural baggage and hence symbolic meaning” (…as…) “different styles of music are often closely associated with different cultural groups” (164). The music thus “evoke(s) the emotions or culture ascribed to these groups” (164), such as a tin whistle for Ireland, a lute for medieval eras or a shamisen for Japan. Because these cultural connections are so powerful, important geographical, historical or mood-setting musical themes with appropriate and strategically limited instrument pallets will significantly enhance the sense of immersion within the fictive world.

With electronic compositions such as those created for many science fiction works, a degree of instrumental restraint can be thought of as an asset to the narrative rather than a restriction on the composer’s creativity. There are literally thousands of software sound generators and sonic manipulators available to the composer, each capable of transforming sounds and instruments in a multitude of ways. It is often extremely tempting, but not always advised, for composers to use any or all of them in a single production. The emotional connections between players and virtual worlds are hard won but they can be effectively maintained by using frequently recurring melodies and a strategically restricted musical instrument and sound palette. This somewhat mandatory unity and repetition means that a well realized score will provide crucial emotive links to the story. On the other hand, a broad palette of disparate music and audio elements is perceived as disjointed to the player, which impairs both the sense of immersion and the narrative continuity of the work.

2.4 Feedback Sounds

Feedback sounds are typically short music flourishes, brief verbal phrases or discrete sound effects that alert, inform, punish or congratulate players. Jørgensen observes that “game audio is not merely ornamental or mood enhancing. It also works as support for gameplay, by providing the player (…with the…) different kinds of information that needs to be comprehended in light of the specific context” (Jørgensen). That is to say, while sound in games must be immersive, emotionally impactful and a narrative enabler, it must also take on an additional didactical responsibility – “to give the player critical game play information” (Stevens and Raybould 279). These usability functions are unique to interactive media and thus set digital games apart from their film and TV cousins.

Off-screen, diegetic feedback sounds play a “crucial role in games (…because of the…) preparatory function that the audio serves, for instance, to alert the player to an upcoming event” (Collins, Game Sound 129,130). Because the cause of these gameplay events is often unseen, the sound “can be used in simple but effective ways in the level design to help navigate players towards areas or objects of interest” (Stevens and Raybould 290). Such ‘attractor’ sounds include those of visually hidden pick-ups and rewards, where “all but the most perverse (…of players…) will be attracted to seek out the source” (290) of the sound. In the stealth action game *Dishonered* (2012) for example, collectables called runes whisper and gently hiss when the player is nearby, alerting them to their presence. Conversely, off-screen warnings or ‘repellor’ sounds such as ticking bombs, grenades, beeping land mines or fluctuating Geiger counter levels are similarly diegetic but reveal zones of the game that need to be avoided. Such sounds rely on the aural modality to “guide (…players…) away from potentially hazardous areas” (291). In many games, unseen enemies whistle, cough, mumble to themselves or talk or each other in order to warn the player of their presence and portend upcoming action or increased stress levels. This banter ceases as soon as all enemies have been defeated, informing the player that it is now safe to wander at will.

An important consideration of the use of sound in games is that of player participation and interactivity. Huiberts states that “in games the user is not a passive observer but an active participant” and that “different emotional effects related to this participation are involved” (Huiberts 39). When compared to linear audio-visual experiences, this ‘acting with agency’ in real time with the sounds of a game world is one of the most rewarding aspects of gameplay. Collins notes that “in games, sound makes players behave in a certain way” and that “although film may act *on* the body, players act *with* games” (Playing with Sound 22). Interactive player action sounds such as gunshots and punches require instantaneity, and it is this very “synchronicity of the (...aural…) response that helps players understand the consequences of their actions” (33). Therefore, although a game may feature both visual and aural feedback, it will “make good use of sound to alert players” or “notify (…them…) that their actions have accomplished something” (Marks and Novak 70, 71), which assures the player “that the action is registered by the system.” (Jørgensen). Such immediate, direct and synchronized aural feedback from player action is immensely gratifying and significantly aids the immersive characteristics of digital games. Therefore, ‘hearing’ a gunshot shot sound in a film is believable and immersive, but ‘causing’ the same sound to occur in a game significantly increases the entertainment value, sense of immersion and suspension of disbelief.

Haptic sensory input from the controller includes physical stick movements, button pushes and even ‘rumble’ sensations for certain important game events but this feedback is lacking in authenticity when compared with the tactile sensations that are experienced in the real world. Stevens and Raybould confirm that, in games, “the amount of feedback that you get through (a controller) is often limited” and that consequently “it is important to have the players interactions confirmed with sound as a proxy for the (…missing…) tactile feedback of the physical world” (281). Player incorporation within the new reality is thus aided by imaginative sound design, which provides “sensory proof of the experience” (Huiberts 47). For example, a real world grenade launcher is a large and weighty piece of cold, hard metal while a game controller is a small, light, plastic object which is completely lacking the tactile sensations of the real thing. The accompanying diegetic audio samples however, typically consist of recordings of the loading and firing of a real grenade launcher, and thus add emotional weight to compensate somewhat for the haptic deficiencies of the controller.

Non-diegetic feedback sounds are used to communicate important game states to the player such as those from the HUD (Heads up Display), menu, game interface or “any button, message, popup, or indicator sound” (Zizza 13). Examples include the successful acquisition of items such as currency, power ups and ammo which liberally sprinkle many virtual landscapes. During intense action scenarios, it is often impossible for the player to keep track of the myriad threats and possibilities that are presented to them by means of on-screen visual feedback alone. Sound effects and voice commands are therefore employed to provide critical game play information to the player in the heat of the moment. Short, non-diegetic music cues are similarly used for feedback purposes as “when you win, an optimistic flourish of sound rewards your efforts (…) with a lot of pomp and circumstance” (Marks 233). While “winning cues will tend to be (...emotionally…) upbeat and composed in a major scale” (233), losing music cues are typically based on atonal sounds or minor key chords.

It was shown in Section 2.3 that, in the name of thematic and narrative unity, musical melodies and their instrumentation palettes are often deliberately restricted in audio-visual media productions. Similarly Stevens and Raybould claim that “your sounds will have more coherence and unity if you link them with the theme of the game world itself” (282) as, if the sonic variations in the sound effects assets are too great, thematic continuity is lost. Consequently, it is important for game sound designers to limit the range of sounds used to some extent, and therefore “the first order of business is to choose a sound palette” (Marks 311). The importance of a sound palette which exists within a stylistically consistent ‘aural theme’ is essential to “ensure that the game sounds have similar qualities” (312). This means that “all audio requires the same character; that is the element that cements it to this, and only this, game experience” (328). The sound design of a game should therefore be unique when compared to other games but familiar within the context of each specific game.

Non-diegetic HUD and interface sounds are typically required to sound identical to each other, rather than merely similar, each time they are instantiated. This repeatability serves “as a kind of conditioned response to punish or reward the player” (Stevens and Raybould 288). If interface, menus and HUD sounds are varied, even in subtle ways, the player could be forced to consider whether these variations were meant to imply additional meanings. There is typically no such sonic variation however, and consequently, “the player is (…only…) concerned with the learned meaning of the sound, rather than any (…additional…) information conveyed in the nature of the sound itself” (288, 289). Therefore, although Chapter 3 illustrates the importance of variation of audio assets in digital games, repetition is specifically required for feedback, HUD and interface sounds.

2.5 Noisic

An intriguingly emotive and immersive relationship often exists between the various sound elements in a digital game. The distinctions between the harmonic, tonal nature of sounds used for musical melody, and the in-harmonic, noisy properties of sound effects, are often considered in terms of discreet disciplines within audio-visual media practice. Music and noise are not always timbrally delineated however, as many real world instruments such as bells, gongs and most metallic sounds contain large numbers of both harmonic and in-harmonic frequencies. Similarly, snare drums, cymbals, hi-hats and many other percussive sounds are, despite being used in the context of a music performance, largely comprised of in-harmonic noise components with no readily identifiable pitch. Sound effects have been shown to be largely noise based but recordings of certain “birds and crickets slowed down many, many octaves (…) eventually become rhythms” (Rona 61), and can therefore function as part of a music composition. Furthermore, the shaker, a popular percussion instrument in many music ensembles, can sound like crashing surf, a sound effect, when significantly pitch shifted down.

A rejection of tonality occurred in the pioneering work of the serial, atonal, and avant-garde music composers of the late 19th and early 20th century. During this period, “music was so rich with chromatics that the tonal continuity was blurred and the sense of the tonic disappeared” (Kaae 80). Moreover, in the latter half of the 20th century a number of experimental sound design concepts such as *music-concrete* and the use of ‘found sound’ in composition emerged. This can be thought of as the birth of sampling as, although tape was typically employed as the medium for recording in that era, the basic temporal principles of sampling still apply. An important aspect of atonal music for game sound is that it is often designed to make “the ears hurt and the body want to escape” (Sonnenschein 106). Hoffert concurs and notes that “atonal music is very likely to be made up of intervals that exhibit tension and dissonance” (40). This highly emotive attribute of audio discord and noise is particularly useful in situations which require that anxiety or fear be evoked in the audience. In other words, virtually every action, science fiction and horror film or game, benefits from atonality and noise at some point in its narrative. Consequently, as “audiences gradually became used to hearing strange dissonances” (Davis 58), these techniques found great favour with composers for films, TV and digital games.

Particularly relevant to this paper is the fact that “many films that blur the distinction between music and sound design tend to be from the science fiction and horror genres” (Collins, The Sound of Music or the Music of Sound). The music scores for science fiction themed films and games comfortably lend themselves to this phenomenon because their instrument palettes are typically laden with numerous sound effects and atonal noises. Whittington describes an early science fiction film example in *Alien* (1979), where “music plays a significant role in establishing the tension of the scene; however the integration of sound (…effects…) is far more disturbing and enervating” (164). Another well-known example is the science fiction classic *Blade Runner* (1982), where “the composition blurs the line between sound effects and music” (188). In this widely studied film, Chion observes that “the audio exists in two, three, or four layers with equal presence – the visual image is just one more layer and not necessarily the primary one. In this vast sonic aquarium, the image is found swimming around like just another fish” (Film A Sound Art 119).

The blurring of the distinctions between music and sound effects is not only timbral, however, as “music is usually outside the fictional (diegetic) world but sound effects are typically contained in that world. (Collins, The Sound of Music or the Music of Sound). As discussed in Section 1.1, this is referred to as trans-diegetic sound by Jørgensen because “when music merges with environmental sounds, it seems to give the impression of being diegetic although technically it is not” (On transdiegetic sounds in computer games 110). As a result, “the overall sonic texture of games can often create an interesting interplay between music and sound effects that blurs the distinction between the two” (Collins, Playing with Sound 3). The music scores of the games discussed in this paper, for example, feature an abundance of futuristic, clangourous and industrial noises, while a number of the in-game sound effects exhibit distinctly musical properties. Jørgensen elaborates by describing the ghostly sound effects which are incorporated into the music score of *Thief 3* (2004): "strictly speaking the ambient sounds of ghosts are extradiegetic (…non-diegetic…) by being part of the music soundtrack, but we could still imagine that these sounds have some sort of diegetic connection since the world of the game does indeed include ghosts" (On Transdiegetic Sounds in Computer Games 112).

This paper proposes the term 'noisic' to refer to this timbral and trans-diegetic muddying of the distinctions between music, sound effects and voice in digital games. Noisic is particularly prevalent in science fiction themed games and "in some games sound effects and music are so intimately linked that they are integrated together in the gameplay" (Collins, Playing with Sound 3). In Chamber 10 of *Portal 2* (2011), for example, the laser receptacle within the diegesis emits a sound effect that is in time and in tune with the non-diegetic music score. Further examples abound, such as the diegetic space ship alarm sound effect in the opening theme music of *Dead Space 3* (2013), which coalesces into part of the non-diegetic melody; or the harsh, sonic rending of an alien 'Reaper' effect in *Mass Effect 3* (2012), which is pitched to the same key as the music score. Furthermore, in several parts of *Mass Effect 3*, fragments of musical tonality from the theme tune fleetingly coexist amongst the sonic murk and electronic abstraction that characterizes the ambient sound design of the game. The organic music score of *Deus EX: Human Revolution* (2011) is so suffused with ambient and noisy tones, and blends so well with the occasionally harmonious environmental sound effects, that it is often impossible to tell them apart.

The noisic phenomenon also extends beyond electronica to many organic, animal and vocal sound effects and, “although speech often has a different function than music, speech-fragments can also be used to form music or can be used as sound effects” (Huiberts 15). Rona further observes that “there is an amazing psychological response to hearing ‘worldly’ sounds played back at slow speeds, harmonized, time stretched or played backwards, (…) it is very disorienting and perfect for setting a strange mood” (61). In *Dishonered* (2012), a metallic sound effect, combined with a noisy vocal phrase, is used to emphasise the player moving into ‘dark vision’ mode. When moving out of this mode, a sonically similar sound effect merges, this time with a backward temporal trajectory, with the environmental sound effects and the music score. This means that the non-diegetic feedback sounds in digital games are often similar in timbre to the diegetic environmental effects track, and similar in pitch the tonal centre of the non-diegetic theme music.

Noisic can be implemented in science fiction digital games without difficulty and merely requires a little cooperation between the composer, voice recordist and sound designer. For example the sound designer could obtain the key of the music from the composer and aim, where possible, to ensure that certain sounds are pitched to that tonality. Conversely, the composer could “treat (…certain…) sound effects like additional instruments in the orchestration” (Hoffert 67). Collins notes that “the more abstract the music the more it becomes a kind of ambient soundscape rather than a ‘song’ in the traditional sense” (The Sound of Music or the Music of Sound). Another technique “that composers use when there are lots of percussive sound effects is to emphasize slower elements in the score (…) letting the sound effects provide some of the rhythm” (Hoffert 67). In addition, the composer or sound designer can obtain some of the voice over material from the dialogue recordist, and, if appropriate, endeavour to find creative approaches for implementing these speech segments within the overall sonic environment. In all cases, the benefits of noisic include an often powerfully menacing soundscape and a more cohesive aural theme for each title.

Chapter 3

Bit Budgets and Technical Conundrums

3.1 Game Durations and Indeterminacy

The extended play times of digital games present sound designers with a unique set of challenges when compared with linear media forms. The average theatrical film release for instance, enjoys an unchangeable duration that will typical last between 90 and 120 minutes. In sharp contrast, the running time of the ‘campaign’ of a game starts at a relatively short seven hours for a puzzle title such as *Portal 2* and hovers around 15 hours for a third person shooter such as *Bioshock Infinite*. This can easily exceed 30 hours for games with branching narratives and/or role playing elements, such as those found in the *Mass Effect,* or *Deus Ex* series. Note that these durations are merely for the campaign or single player story component of the game, and that the time spent by a player in the online multiplayer mode of these titles often extends into many months. The most extreme example of lengthy game play is found in Massively Multiplayer Online Role Playing Games (MMORPG) such as *Starcraft 2* (2010), which can engage the player for years. In contrast, the linear and highly predictable nature of film and TV narratives require a smaller number of predetermined, non-varying sound assets, albeit of an extremely high quality.

To complicate matters further, Collins observes that, “unlike the common uses (of sound) in a film or television programme, a game would typically be played over and over, for hours and hours on end" (Collins, From Pac-Man to Pop Music 1). Many game narratives include a branching storyline, "in which there are many possible paths and endings" (Collins, Game Sound 142) and, as such, virtually demand to be replayed a number of times to experience the additional narrative permutations thereof. Moreover, many games offer at least one new difficulty level or play mode on completion, which further encourages replay. Games with RPG elements feature an enormous variety of outcomes, many of which are determined by character choice, player statistics and character upgrades or development; all of which are variable, depending on the player. As a result, "the impact of the overall branching narrative structure on audio planning (…in games…) is significant" (Collins, Game Sound 143). Non-linear narratives with branching paths therefore require far more audio content than linear structures of a similar quality.

Apart from longer running times and branching narratives, "the most significant problem facing game composers is the non-linear basis of games in general" (Collins, Game Sound 142). Stevens and Raybould note that "the greater the interactivity there is in your game, the greater the number of possible outcomes and therefore the greater the number of possible sounds" (35). Therefore the amount of sound required will vary arbitrarily depending on the skill of the player, who may get stuck for longer in certain levels than others or simply choose to linger. As Marks observes, "game players are not predictable. We don't know when they will walk, run, hide, enter a new room, meet the bad guy, draw their weapon, or do any one of the other hundred possible actions that can happen during a game" (234). This indeterminacy of gamer playing style therefore considerably complicates game audio development.

In the real world, the sounds of many material objects subtly vary in timbre, duration and amplitude each time they are instantiated. These variations are barely perceptible, but nonetheless provide important clues as to the veracity of the sound sources. A noticeably repeated diegetic sample, on the other hand, compels the player to acknowledge the fact that this is merely a game, replete with all of the multifaceted technical issues of modern game audio. Stevens and Raybould observe that “the chief criticism of game audio over the years has been its repetitive nature” (43). This principle does not apply to feedback sounds which, as discussed in Section 2.4, most often require familiarity in order for their usability functions to remain effective. Diegetic, game world samples, however, require variety because “nothing breaks immersion in a game more than hearing the exact same sounds/samples being used repeatedly, as few sounds in the natural world repeat in this way” (35). The practical outcome of this is that, in modern game design, a single high quality audio sample for each visual object is rarely adequate. In order to facilitate the illusion of verisimilitude of sound assets, several samples, each one subtly different from the other, are required for every discrete game sound. Subsequently, when the sample is triggered by the game or the player at runtime, the game engine will randomly choose a different version of it for each instantiation. It is clear therefore, that digital games require an enormous amount of variation in audio material to offset boredom and prevent noticeable repetition during extended periods of game play. This is particularly important when the player remains in one area of the game world for a while, whether to explore for collectables such as loot or treasure, or when trying to solve a particularly difficult puzzle or similar game challenge.

3.2 Storage Technologies and File Sizes

A crucial dilemma in game sound practice is where to place the multitude of audio assets so that they are instantaneously available. There are a number of technologies available for storing and delivering game titles to consumers but many of them are mechanical by design and therefore not suitable for interactive audio purposes. By mechanical I mean that the storage technology is distinguished by moving parts as opposed to solid state circuits. Hardware game consoles and PCs often use motorized DVD and Blu-ray optical disc players as a delivery medium, but these playback technologies are currently the slowest in terms of access times. Much of the game content is therefore pre-loaded from an optical disc or the internet to a magnetic hard drive, either when installing the game or during the load time for each level. Magnetic hard drives are also mechanically based and, while significantly faster access times than optical discs are possible, this technology still does not achieve the instantaneity required for real time audio interactivity. Consequently, on any gaming platform, the only technology available which is fast enough to provide immediate audio playback on demand is RAM, into which all interactive samples must be pre-loaded.

RAM is the speediest storage medium currently available, but it is relatively expensive, and therefore a strictly limited resource on game hardware devices. In addition, the sound assets must compete with high resolution graphics for this same storage space. Andy Farnell refers to the use of audio samples in games as the ‘data model’, as this is a particularly data intensive practice (217). The increased audio asset requirements of digital games, complicated by the limited amount of RAM available at runtime, is therefore yet another dichotomy which game sound practitioners must deal with.

Because it is rarely possible to increase the amount RAM, many game audio strategies involve instead the reduction of the size of the audio files stored in it. Audio which is referred to as ‘CD quality’, ‘PCM’ or ‘uncompressed audio’ is sampled at a rate of 44.1 kHz, and quantized to a 16 bit resolution. Note that, although somewhat related, the use of the term ‘sampled’ in this case refers to the digitization process, as opposed to a complete audio file which responds to triggers as explained in Section 1.1. PCM audio typically requires around 10MB of storage space for each stereo minute of playback time, and consequently file size is always a prime consideration for game audio practitioners.

The audio files used in early digital games limited their file sizes by either reducing the sample rate to a 22.050 kHz or 11.025 kHz, or truncating the resolution to eight bits per sample. While effective from a bit budget perspective, such crude file size reduction methods result in perceivably lower quality sounds denigrated by distortion artefacts. On the other hand, “the higher the sample rate and resolution, the larger the audio files – so size and quality concerns are significant factors” (Marks and Novak 20). PCM audio is an extremely inefficient digital coding process, which accurately samples all sound at the same fixed resolution, no matter how unimportant certain aspects of it are. One particularly wasteful property of PCM audio for example, is that “silence takes the same amount of memory as sound” (Stevens and Raybould 37).

During the second half of the 1990s, compressed audio file formats were added to the list of basic memory saving techniques. Modern data reduced audio files use perceptual coding algorithms, which shuffle the bits around to place more of them where they are most required and use fewer for the less important aspects of the sound. The popular mp3 and similar formats for example, can “achieve compression (…ratios…) of up to 10:1” (Stevens and Raybould 38) or around ten times smaller than its PCM counterpart, with a relatively insignificant perceivable loss of audio quality. Contrast this with the mere 50% file size reduction that halving the bit depth or sample rate accomplishes, and with a considerably greater loss of sound quality, and the reasons why compressed audio file formats have become standard game audio practice become apparent.

Another advantage of compressed audio file formats in digital games is that they are more flexible than PCM audio because the amount of compression or data reduction can be varied to suit the occasion. Higher bit rates provide improved quality while lower bit rates suffer a loss of in perceived quality but with a correspondingly smaller file size. This is an essential attribute for game audio as it means that low level background sound effects can be coded at a lower resolution, while more important audio assets such as dialogue will make use of the increased quality of higher bit rates. For example, excessive data reduction “degrades the higher frequency band that is very noticeable for sounds with most of their focus on the higher range, such as shell casings (…and…) bells” (Marks and Novak 97). These sounds with a bright timbre will thus require more bits and therefore larger file sizes than background ambiences with a relatively dull spectral content such as distant thunder.

There are now a number of compression formats in existence and “each type of file uses its own unique brand of compression and sound quality” (Marks 213). Microsoft for instance, uses the proprietary XMA format for its Xbox hardware, which in turn is built upon the Windows based WMA architecture. The mp3 file type is patent protected and licensing fees must be paid to the numerous rights holders for its use. Conversely, Ogg or Ogg Vorbis files are ubiquitous in game audio practice due to their free, open and non-proprietary nature. In addition, Ogg is an extremely efficient audio codec and one minute of 44.1 kHz, 16 bit stereo .ogg at a bit rate of 128 kbits/sec results in a file that is only 186kb in size (Marks 214). Game audio practitioners “will often use both compressed and uncompressed formats depending on the priority (…the sound…) has within each segment of the game” (Marks and Novak 159). In addition, “considerations such as the targeted game platform, type of storage medium, and even cost of packaging can have an impact on how much space is available to the developer” (20). No matter which data reduction scheme is used, however, “the big trade-off” (Marks 210) of extreme compression is always sound quality, and consequently a degree of give and take between quality and file size is unavoidable.

The conflicting requirements of quality, variety and file size in digital games are referred to by Stevens and Raybould as the “Triangle of Compromise” or the “Triangle of Pain!” (36).

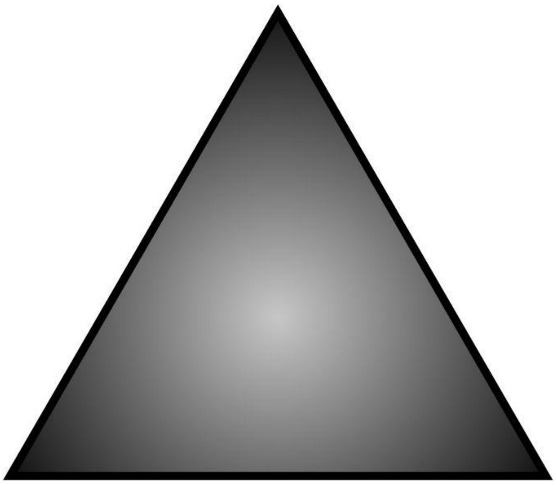

Figure 3. The Triangle of Pain as illustrated by Stevens and Raybould

Figure 3 clearly illustrates this quandary. The more variation and quality of audio assets that is required, the more memory will be consumed. Conversely, when, “concessions are often made in order to squeeze everything in, decreasing the quality of the audio is typically considered the lesser of the evils” (Marks and Novak 20). It is often music quality which suffers in this regard, “and even today seems to be the first set of assets a developer degrades when faced with an obstacle” (158). Quality considerations are particular important for console and PC games which may be played over high end home theatre systems. On the other hand, “speakers for portable consoles (…and mobile phones…) do not have a great frequency range so there would be no point” (Stevens and Raybould 42). No matter which platform is used however, RAM is always in short supply and “typically, sound is (…only…) allowed to occupy only 5 to 10 per cent of system resources” (Collins, Game Sound 82). It is therefore obvious that, in order to cram believable sounds into the cramped confines of the available system RAM, additional memory saving strategies are required.

3.3 Looping

Despite the use of audio data compression, the creation of dozens of hours of mutable audio assets to cater for long gameplay durations is rarely financially or technically viable for game developers. One solution to this dilemma is to loop the audio files whose required durations are indeterminate. However, certain sounds are easier to loop than others and “looping is good for (…) sounds that are, by their very nature, constant or repetitive” (Stevens and Raybould 24), such as sections of music or background ambiences. Loops are thus samples which are designed to play indefinitely and because “looping extends the duration of sampled sound” (Roads 121), loops are particularly effective at making a little go a long way. Looping audio files allows for “continuous accompaniment for a few seconds to infinity” and “will play on (…continuously…) until a transition is triggered” (Marks and Novak 151). In addition, short loops can “quickly load into RAM and play repeatedly without placing further demands on the processing pipeline” (Marks 235). Looping has therefore become a widely used bit budgeting technique for game audio practitioners.

Preparing an audio file for looping involves creating “a head and tail which match each other seamlessly” (Hoffert 26), so that the loop will repeat with no noticeable glitch or hiccup at the boundary. In addition, and “to reduce the likelihood a player will notice any repetition, the loop should be as long as possible” (Marks and Novak

151). Although music melodies “provide deep emotional connections to game characters”, they are “easily memorized and become irritating when repeated over long gameplay durations” (Stevens and Raybould 43). Marks and Novak elaborate: “music created for (…looping…) is generally simple and open” (153). On the other hand, while “the use of ambient music with no discernible melody is one solution to this problem” (43), such music does not evoke the same feelings of identification with the game world. Westerners are used to linear music with a “goal oriented, directed nature” but the lack of directedness of open music forms leads to “the elimination of the dramatic curve” (Collins, Game Sound 158). Collins is thus describing the inescapable temporal trajectory of structured music, which is a powerful emotive and narrative enabler. From a science fiction perspective however, many short, noisy and barely melodic ambiences are effortlessly loopable, as a result of their lacking a readily identifiable musical theme. Thus an additional benefit of sounds which fall in the noisic category is a tiny memory footprint.

Another problem which requires careful consideration, especially in the case of music loops, is the unpredictable, branching narrative structure of digital games in general. Music loops should not only repeat seamlessly, they must also be able to smoothly transition between any other loops which may precede or follow them. Collins elaborates: “Since the music may have to transition at any point in time, a potentially infinite number of cues may be needed” (Game Sound 146). In addition, “predicting how these cues may have to connect with other cues (…) can be very difficult” (145). Rhythmic music in particular is fraught with temporal implementation obstacles which are specific to interactive media. The instant that the piece responds to an interactive trigger “could occur at any point in the music, but normally the response is set to occur immediately, at the next beat (…or…) at the next bar" (Kaae 84). If the music switches immediately, it means that there is no timing reference to its rhythmic structure. Such immediacy of response will be truly interactive but musically incorrect and therefore potentially distracting. A musical change on the next beat or bar however, requires the trigger to wait for a fixed period of time. Van Geelen elaborates: "to make the gameplay fit the rhythm of the music, small delays in reaction to the player's input could be used to make the player's character do his/her thing exactly on the beat of the music" (94). However, although such delays make rhythmic and musical sense, they destroy the instantaneity, and therefore the interactivity of the sonic response to player actions.

Looped ambient or environmental sound effects can also be troublesome as "the player will quickly pick up on any distinct sound that repeats" (Collins, Game Sound 93). For example, if an outdoor ambience contains frog croaks, bird chirps and barking dogs, the order in which these conspicuous audio events occur will become apparent over numerous recurrences of the same file. In addition, the high quality of samples "only helps to expose the fakery as soon as the same sound plays for a second time" (Farnell, Behaviour, Structure and Causality in Procedural Audio 323) Collins observes that "the persistent practice of looping is particularly illustrative of the tensions between technology and aesthetic" (Game Sound 34) in digital games.

3.4 Sample Remixing and Re-sequencing Strategies

The practice of looping is a relatively crude attempt to compel temporarily stubborn samples to behave in a more flexible manner, but there are many other temporal strategies available to game audio developers. For instance “you can break up the looping feel by having another (…similar…) sound of a slightly different length play at the same time” (Stevens and Raybould 24). These two samples have different durations and therefore their loop boundaries or repeat points will occur at a different moment in their cycle. This means that, while looping, they will continuously “drift out of time with one another, so the resulting combination of the two sounds will be different every time” (24). This is referred to as ‘phasing loops’ by Stevens and Raybould as “the two loops phase in and out of time” to “create new rhythms and combinations” (25). Phasing is a particularly useful strategy for ambient sound effects to prevent the noticeable repetition of certain sonic elements that would otherwise occur. The advantages of this technique are that it requires relatively little RAM and it takes a considerable amount of time for the two loops to drift back into sync with each other.

Another remixing option, alternately known as “vertical re-orchestration” (Marks and Novak 147), “vertical layering” (Stevens and Raybould 213)’ or “parallel composing” (Van Geelen 97), is often used to create the illusion of a more temporally flexible game music score. This is a technique in which the arrangement or mix of each piece of music is split into a number of individual but musically matching ‘separates’, ‘tracks’ or ‘stems’. All of the separates start at the same time and play simultaneously for the duration of the level, but not all of them are audible throughout. The game engine is “used to mute or unmute certain stems in order to produce different versions of the full arrangement” (Stevens and Raybould 215). For example, a slow and gentle drone could play continuously in an exploratory area of a level. If enemies appear, an accretion of music layers, each with greater amounts of temporal activity will be mixed into the score. But, because the original drone is a mutual element of both activities, there is less of a noticeable disconnect when the music changes to ramp the intensity up or down (213). An additional advantage of this strategy is that “smooth transitions in volume” of the layers can be implemented “in real time (…at runtime…) in response to gameplay variables” (219). This means that music transitions which are faded in and out at runtime, will be less obvious than the sudden switching tracks, and will therefore not disrupt the suspension of disbelief.

Music can also be manipulated horizontally or temporally, in which it is written as “relatively short blocks”, which can “react (…) quickly to game events while maintaining a musical continuity” (Stevens and Raybould 193). This strategy is referred to variously as “horizontal re-sequencing” (Marks and Novak 146), or “horizontal switching” (Stevens and Raybould 84). Either way, the technique implies that “segments of musical material are constantly chosen at random and put together (…at runtime…) to form a piece of music” (Kaae 84). In other words, music which is already mixed and finalized vertically into a single score is divided temporally into musical phrases, melodies or motifs. These pieces of audio material are “shuffled (…horizontally…) according the in-game movements of the player” (Marks and Novak 146). Such sonic segments are inescapably temporalized but, the shorter they are in length, the faster they can adapt to game states and the more interactive they will appear to be. Volume crossfades can be inserted at the transition points to smooth out the changeover from one segment to another, or very short music cues called ‘stingers’ can be added at the temporal boundaries to mask the transition completely (146). However, because the final music sequences are no longer linearly chosen by the composer or programmer, “there is no single musical text and (…an additional…) author is, to some extent, the player/listener” (Collins, Game Sound 106). The notion of ownership of such works therefore becomes somewhat nebulous and the copyright implications thereof require further debate.

Concatenated sound is similar to horizontal switching or transitioning but most often implies chaining a group of very short segments of audio together (Stevens and Raybould 69). These sonic snippets are typically 100 – 500 milliseconds in duration and are sequenced into chains whose order can be randomized at run time. A concatenated sampling technique commonly used in games is to “break a sound into its component parts and to use scripting and other implementation techniques to describe an algorithmic method of playback” (Paul 137). Concatenation is especially useful for creating endless, evolving ambient sound effects such as crackling fire, footsteps, jungle noises or background radio chatter (70). For example, when creating concatenated footstep sounds, only a few audio recordings are required for each surface, and these are split into left and right foot variations. These footfalls are then further segmented into individual heel and toe components, to produce a number of extremely short samples. When such audio files are randomized and chained together appropriately, the footstep sounds appear to subtly vary on each step.

Granular sampling technologies can be thought of as the most severe form of concatenation as they sequence thousands of tiny slices, or grains, of sampled audio. Grains are minute specks of sonic substance “of very short duration: 10–100 milliseconds” (Russ 294), or “near the (…durational…) threshold of human hearing” (Kuehnl). In other words sampled grains are so brief that they would be virtually inaudible without the simultaneous presence of adjacent grains in the chain, and are thus “discerned as one large sonic mass” (Kuehnl). Each granular audio flake is individually pared down and amplitude enveloped to create a short volume fade-in and fade-out at the temporal extremities of the grain. This reduces noticable amplitude of clicks and pops at the boundaries but also allows the grains to be overlapped (crossfaded) and looped individually. The “atomistic view of sound as ‘particles’” (Roads 168), “applies to the microstructure of sound” (Farnell 89), and accordingly implies microscopic (or micro-sonic) control of articulated audio shavings. For example, grains are temporally sequenced in interesting ways at runtime and can be repeated, omitted or re-ordered.

The advantage of granular techniques is that one sample, with a low memory footprint, can be sliced and granularly randomized so that “it is never heard in the same way twice” (137). Granular techniques therefore partially solve three game audio dichotomies in one fell swoop: they retain some of the veracity of recordings of tangible sounds, they occupy a relatively insignificant amount of RAM and they are not repetitive. In addition, these grains can be scattered to create what is known as a ‘cloud’, “a precision spay jet for sound, where each dot in the spray is a sonic grain” (Roads 176). This “amorphous soundscape” (Collins, Game Sound) allows virtually limitless temporal combinations and permutations of the original sample. Grain clouds are therefore capable of providing endless, evocative and non-repetitive audio ambiences for digital games, but are using source content that is still rooted in reality.

Granular synthesis is not a game audio panacea however, as it is often only truly viable for certain ambient sound effects or atonal music. This means that natural speech and acoustic music scores can seldom be convincingly granulated unless the resulting granulation artefacts are acceptable or specifically required for the game. Nevertheless, the discordant and noisy sonic by-products of granulation are often particularly useful for sound effects ambiences in science fiction themed digital games. For example, continuous, airy and in-harmonic drones often act as convincing substitutes for the sound of space, or on-going space flight. In some cases, granulated melodic fragments will crumble under the weight of the additional sonic artefacts to create another example of noisic, with all of its previously discussed benefits to game audio practice.

3.5 Runtime Parameter Manipulation

The creation of quality source samples is only part of the digital game immersion equation. Game audio in its entirety can be described as “a system of sound and music elements that interact in real time with game events” (Stevens and Raybould xvii). This means that, “you are not dealing with ‘an event = a sound’, instead, you are constructing a model or system to deliver dynamic content from the static sound waves” (333). In other words, a game audio asset is not merely a triggered or looped sample; it is the sample together with the myriad real time parameters that determine how it is played back interactively within the game. The horizontal looping strategies that were discussed earlier in this chapter involve the temporal shuffling of sampled audio elements at run time to create the illusion of interactivity. This illusion can be further enhanced, without significantly impacting the memory footprint, by the real time parameterization of the amplitude, timbre, pitch and spatialization properties of samples. This is typically referred to as DSP (Digital Signal Processing) and is “one of the most important advances in sound technology on next-generation consoles” (Collins, Game Sound 95). DSP requires additional processing power rather than RAM storage and can be thought of as automated mixing at runtime.

Mixing is concerned with the effective prioritization of multiple layers of sound effects, voice and music elements in order to create believable virtual worlds. This concept is often referred to as the ‘final mix’ in linear forms of audio-visual media but such post-production audio mixing does not exist in digital games. Collins notes that “sound and music in film are often closely tied to the edit” (Game Sound 128) and this means that the timelines for both the audio track and the moving images of a film are synchronised with each other. In other words, mixing for linear audio-visual media is not as complex because, “film audio is generally a post-production activity that takes place after the film has been edited and the visuals locked” (89) in time. However, because “game timings and interactions are unpredictable, (…) mixing must essentially be accomplished in real time” (Collins, From Pac-Man to Pop Music 9). To put this another way, the process of mixing a feature film soundtrack can take many months, as each audio element is carefully adjusted in amplitude, timbre, and spatiality, and precisely positioned in time. Game audio practitioners, on the other hand, are expected to create an implementation system which performs without warning, a mix of similar quality, and in a fraction of a second.

Mixing in all audio-visual media does not attempt to recreate the sound of real life as this is neither possible nor desirable. Marks and Novak notes that, “the human mind contains a subconscious mechanism that filters out unimportant background noise allowing only significant and life-impacting cues to be heard” (236). This psycho-acoustic phenomenon is often referred to as the ‘cocktail party effect’, because at such events, we can selectively tune in to whichever conversation we find most interesting, despite the chaotic nature of the total soundscape. Within the context of digital games, however, the programmer can only try to simulate this phenomenon by foregrounding important sounds at the expense of others. The mix thus prioritizes entertainment value over verisimilitude or, as Stevens and Raybould put it, “we don’t mix to make sure that everything is heard, (…) we mix to ensure the sounds we want to be heard are heard” (295). In addition, and as discussed in Section 2.2, real life soundscapes are not as exciting, or as entertaining, as the hyper-stylised sounds used in mainstream audio-visual media productions.

The perceived distance and direction of sound sources from the player/character is commonly referred to as perspective in games but is “fundamentally different from perspective in films” (Collins, Playing with Sound 49). Unlike film, “the player usually has some control over both the visual camera and the auditory perspective” (49) in games, and is, in essence, mixing and panning the final audio track. Panning is the directional placement of audio assets within a three dimensional sonic space but occurs in games, “in a more complex fashion due to the player continuously changing perspective" (49) In other words, "this type of panning is more dynamic, since constant evaluation within the programming is made to determine the objects' position in relation to the view of the character" (Marks and Novak 251). This means that the player is in control of the final playback positions and amplitude of many of the audio assets and is essentially mixing the game at runtime.

Another important mixing and positioning tool is the manipulation of timbre, using what is variously referred to as a filter, tone control or equalizer. Stevens and Raybould observe that "air acts a little like a low pass filter, removing the higher frequencies over distance" (Stevens and Raybould 104). This means that loud, bright sounds appear close by, and quiet, dull sounds further away (Belton 194). This is effect is simulated by programming both the timbre (a high frequency roll-off) and amplitude, to diminish dynamically in accordance with a quantifiable distance between the player and sound source. On the other hand, the high frequencies of sounds are often emphasized, to bring them closer in perspective or to create "a feeling of temporal acceleration" (Chion, Audio Vision: Sound on Screen 98). This enhanced "definition" of the timbre of a sound "lends itself to a more lively, spasmodic, rapid, alert mode of listening" (98), and engenders additional concentration and immersion of the player within the virtual world. In addition, timbral manipulation can be used to elicit interesting science fiction or fantasy effects for audiences. First person shooter, *Crysis 3* (2013), for example, manipulates the timbre by removing high frequencies whenever 'cloak' mode is engaged. Therefore, whenever the player is shrouded in invisibility, he or she is also shrouded sonically, greatly facilitating their suspension of disbelief.

Many spatial effects are likewise created and manipulated at runtime. Natural ‘*echo*’ occurs as the distinct repeat or reflection of a sound off of a large surface such as a building or a cliff face, and thus typically occurs over considerable distance outdoors. Real world ‘*reverb*’, on the other hand, is the thousands of short reflections which occur when sound is emitted and contained within a six sided acoustic space indoors. Both of these spatial effects are easily simulated in software using artificial echo and reverb devices. Such “digital signal processing effects that simulate a space” (Collins, Playing with Sound 55) , can add further variety to a limited number of samples. Footsteps, for example, do not require additional samples for when a player walks into a tunnel, cave or cathedral; as spatializing effects are simply added to the original ‘dry’ recordings in real time using game engine DSP.

There is more to the representation of virtual acoustic spaces than allowing simple indoor/outdoor variations of sounds through the use of reverb and echo. Marks and Novak observe that “sounds are not always heard directly from their source, and they instead may be muffled by a wall or blocked by other objects” (249). This muffling is typically referred to as occlusion and implies a drop in both volume and high frequencies of the direct sound, with a corresponding increase in the amplitude of the spatial effects. Such “occlusion and diffraction effects (…occur…) where objects in the space obstruct and alter the sound” (Collins, Playing with Sound 55) and must be implemented at runtime in order to maintain the sense of immersion.

In the critically acclaimed stealth action title, *Deus:Ex Human Revolution* (2011), a simple error in occlusion effects destroys the otherwise plausible sense of audio ambience that occurs within the Detroit Police Department level of the game. While exploring the admin offices of this level, a stationary civilian NPC (Non Playable Character) repeatedly wails “help me please” and many aspects of this specific voice over are well implemented. The acting is expressive and convincing, the recording and playback quality ensures clearly articulated speech; and, with the player having recently killed all police officers inside the station, the script perfectly suits the situation. However, the believable veil of audio immersion is largely upended by the spatial characteristics of the dialogue which consistently sounds as if the NPC is standing only a few meters from the player, even as the player moves its avatar a considerable distance away, around corners of the building, into other offices and behind large pieces of furniture. The dialogue implementation of this particular character does not take into account the sonic implications of amplitude and high frequency loss over distance, or the occlusion effects caused by obstructed sound.

Runtime parameter manipulation can also be applied to the pitch, direction and envelope properties of samples to create additional variety from a limited pool of samples. Stevens and Raybould elaborate: “when memory is important (and it’s always important), then try to use the same sound in different ways by changing the pitch at which it plays back within the game engine” (43). In addition, many sounds, when radically pitched up or down, take on entirely new personalities with no resemblance to the original recording. Certain percussive instruments, for example, sound like thunder when pitched down several octaves, and the scraping of bricks can be easily transmogrified into the ominous sound of a grave-stone opening. Envelopes too can be randomized at runtime and “the use of randomized fade-in and fade-out lengths (…) allow you to alter the volume envelope of a sound to get a huge range of variation from one sound in memory” (77). Moreover, a single sound can be played backwards or forwards, used as an impact effect or subtly looped and cross-faded in the background – all of which sound completely different to each other.

An additional consideration in game audio is that the number of voices that can play back simultaneously is restricted, not only by the RAM available, but also the game audio hardware capabilities. This voice ceiling is often reached, and the audio engine must then intelligently handle any loss of sound elements that may occur. Moreover, “the audio programmer, or the audio team via scripting, will tag specific sounds that will always play, and those that will be muted as more sounds enter the picture” (Marks and Novak 239). The mute button is therefore one of the most powerful mixing tools as it decides “the number of (…simultaneous…) sound sources, (…) and which can be lost or culled” (Stevens and Raybould 295). In other words, if we accept that some voices will certainly be lost due to the hardware limitations, the programmer must control this loss so that it is less noticeable than it would be if their disappearance from the soundscape was left to chance.

A significant number of game hardware technical limitations have thus been partially circumvented as a result of the myriad innovative implementation strategies that have evolved in mainstream game audio practice. In addition, mobile gaming platforms continue to rise in status as a viable option for digital games, but such devices are typically a technological generation or two behind their better specified console brethren. Consequently, many of the audio implementation strategies discussed in this paper will no doubt become increasingly relevant for mobile gaming over the next few years.

Chapter 4

Predictions and Conclusion

4.1 Procedural Audio

The near ubiquitous use of corporeally eager, temporally pig-headed and uncompromisingly data-centric samples as the primary sound source in digital games, has, in many ways, forced game audio developers to wrestle with the laws of physics. Therefore the current methodologies employed in digital games are to “collect as much data as possible in finished form and assemble it like a jigsaw puzzle, using a hammer where necessary, into a product” (Obiwannabe). In addition, the laborious practice of “curating and matching audio data to depicted circumstances” (Farnell, Behaviour, Structure and Causality in Procedural Audio 313) means that asset management has become a serious problem.

The only viable alternative to samples that has been proposed so far is a return to synthesizing sound from within the game engine at runtime. The synthetic sounds which preceded the domination of samples lacked verisimilitude and only a few synthesis techniques were available at the time. However, synthesis techniques have improved exponentially in the 20 years since they lost the first instalment of the game audio battle, and a number of game audio theorists feel that they are worth a second look. Therefore a “promising possibility for sound effects (…and only sound effects at this stage…) hinges on the work being done towards their generation in real time and on the fly” (Marks 11). To that end “current research is realizing the potential of procedural audio for generating sound effects in computer games" (Mullan 340). Procedural implies sound as a process rather than an event, and "the product is 'potential sound' rather than particular audio data" (Farnell, Behaviour, Structure and Causality in Procedural Audio 314). In other words "the use of the word 'sound' (…sample…) implies that it has already realized (concretized) as a signal while 'procedural audio' is still code" (315). The advantages of procedural audio are that it is RAM free, temporally malleable and infinitely reusable. In addition, there are a number of game audio engines currently available which allow the production of certain synthetically generated sound effects at runtime.

A significant disadvantage of procedural audio, and one which seems to receive the least attention, is that it is currently only able to provide convincing realizations of a few sound effects such as weather, water, fire and footsteps. In other words natural speech and acoustic music scores cannot be affectively and believably synthesised using modern runtime technologies. Speech synthesis techniques have been in development for decades, but still provide nothing close to the emotive qualities that a real voice actor is capable of. For example, no current speech synthesizer can convincingly deliver the humour, sarcasm, anger or the myriad other human vocal emotions that are required by a game narrative. Similarly, pre-scored and pre-mixed music recorded in a professional recording studio is typically more powerful, emotive and engaging than any real time synthetic renditions thereof. The sonic magnificence of a real orchestra when accompanying visuals will thus remain a cultural necessity for many game genres and even electronic music can currently be more dramatically realised in a studio than in code. This means that at least two thirds of all game audio sources must be sampled, and this situation is unlikely to change for the foreseeable future.

4.2 Predictions

Samples are here to stay, at least until audio synthesis techniques have evolved sufficiently. Sample playback remains a natural fit to the linear audio production methods used by all composers, sound engineers and sound designers. Because music and speech can only ever be linearly presented using current technologies, most recording devices in use throughout the audio-visual media industry are based on the unremittingly linear Digital Audio Workstation (DAW). This philosophical chasm between the linear functioning and tool set of the majority of audio practitioners and the non-linear requirements of game sound implementation is gargantuan. As a result, Matt Piersall (in Marks and Novak) notes that “the tendency of sound designers and composers is to create their assets in a way disconnected from the final product”. (61) This means that the creators of audio assets, and the implementers of them, are not always on the same team, literally and figuratively. The division of labour in game audio thus presents a conundrum in that, while specialization in each field implies greater proficiency and, by extension, sonic quality; co-operation commonly provides the most immersive and engaging experience for the player.

A number of partial solutions to the almost mutually exclusive game audio skill set problem are emerging. Game audio programmers can work in acoustically treated studios with mastering grade monitor speakers, which enable a professional audio mix-down environment. The more creative or musical the background of the programmer, the more likely he or she will be able to use such high end facilities to create believable and affective runtime soundscapes. Similarly, the more the composers and creators of audio content understand the rudiments of programming or scripting, the more likely they will be able to relate to the punishingly limited audio options available at runtime. Game audio middleware platforms such as FMOD have introduced software components and interfaces that are more linear and DAW like, and this may be another step in the right direction. Ultimately, a mutually coextensive production and implementation software platform that serves the needs of both game audio creators and programmers will help alleviate a significant number of these difficulties.

All of the implementation techniques discussed in this paper will remain relevant for the foreseeable future. Some theorists have argued that RAM availability, or the lack thereof, is the most significant problem facing game audio practitioners, but this is not the case. An infinite amount of RAM would allow more uncompressed audio assets to be available interactively, thereby providing an improvement in audio quality and more interactive possibilities. However, no matter how much RAM is eventually available to gaming hardware devices of the future; it will never be logistically practical or economically viable to create hundreds of hours' worth of audio assets to fill it with. In other words, the problems of extended gameplay durations, branching of game narratives and player unpredictability are not going anywhere. If anything, more branching of narratives, further use of multiple potential endings and longer gameplay durations are on the cards for the future of digital games. Therefore, even when the RAM limitation conundrum is completely eliminated, every implementation strategy in current use would still be central to game audio practice.

4.3 Conclusion

This paper set out to illustrate the fundamental intractability of using recorded sound in interactive contexts. The game audio strategies discussed, therefore, largely involve wide scale temporal fakery and myriad ham-fisted efforts to shoehorn static samples into interactive modes of operation, while simultaneously working with just a fraction of the storage resources enjoyed by linear audio-visual media forms. To put this metaphorically, game audio practitioners have spent the last two decades trying to knock a square peg into a round hole and cram a quart into a pint pot. This pint pot (RAM) will undoubtedly evolve into a gallon jug in time, but will ultimately solve very few game audio dichotomies. The square peg (the steadfast and inflexible temporal qualities of most real world sounds) will continue to remain a game audio conundrum. Therefore, if we agree on the term ‘audio’ to imply all natural and synthesized aural experiences (including speech and music), then, as technology currently stands, interactive audio is a myth for the majority of sounds required by game practitioners. Even if we recognise that certain synthesized sound effects can be made to behave interactively, or acknowledge that brief audio samples are largely immune to temporal problems, one fact will remain indisputable – game audio is not for sissies!

Works Cited

Biography

Altman, Rick. *Sound Theory/Sound Practice*. Routledge, 1992. Print.

Altman, Rick. "The Evolution of Sound Technology." Weis, Eilsabeth and John Belton. *Film Sound: Theory and Practice*. New York: Columbia University Press, 1985.

Arsenault, Domenic. "Video Game Genre, Evolution and Innovation." *Eludamos. Journal for Computer Game Culture, Vol 3, No 2* (2009). Web - http://www.eludamos.org/index.php/eludamos/article/viewArticle/65/125.

Ashline, William L. *Vague Terrain: Digital Art/Culture/Techonololgy*. n.d. 23 December 2013.

Beauchamp, Robin. *Designing Sound for Animation*. Burlington: Focal Press, 2005. Print.

Belton, Elizabeth Wies and John. *Film Sound: Theory and Practice*. New York: Columbia University Press , 1985. Print.

Bordwell, David and Kristin Thompson. *Film Art:An Introduction*. New York: McGraw-hill, 2008.

Bordwell, David and Thompson. "Fundemental Aesthetics of Sound in the Cinema." Weis, Elisabeth and John Belton. *Film Sound: Theory and Practice*. New York: Columbia University Press, 1985. 181.

Boyd, Justin. *Marginallia*. n.d. 1 January 2013.

Breitlauch, Linda. "The Auditive Effects in Movies and Computergames." *The Art of Non – Music* n.d. Web - http://gdcvault.com/play/1015367/The-Art-of-Non-Music .

Burch, Nöel. "On the Structural Use of Sound." Weis, Elisabeth and John Belton. *Film Sound: Theory and Practice*. New York: Columbia University Press, 1985. 200.

Chandler, Daniel. *Semiotics: the basics*. New York: Routledge, 2007.

Chion, Michel. *Audio Vision: Sound on Screen*. New York Chichester: Columbia University Press, 1994.

—. *Film A Sound Art*. New York: Columbia University press, 2009. Print.

Clearwater, David A. “What Defines Video Game Genre? Thinking about Genre Study after the Great Divide.” *The Journal of the Canadian Game Studies Association* (n.d.). Web.

Collins, Karen. *From Pac-Man to Pop Music*. Aldershot: Ashgate Publishing, 2008. Print.

—. *Game Sound*. Cambridge: MIT Press, 2008. Print.

—. *Playing with Sound*. Cambridge, Massachusetts: The MIT Press, 2013.

—. *The Sound of Music or the Music of Sound*. 2013. 31 January 2014. <http://designingsound.org/2013/03/the-sound-of-music-or-the-music-of-sound/>.

Davis, Richard. *Complete Gude to Film Scoring*. Boston: Berklee Press, 1999.

Everest, F. Alton. *Master Handbook of Acoustics*. New York: McGraw-Hill, 2001.

Farnell, Andy. “Behaviour, Structure and Causality in Procedural Audio.” Grimshaw, Mark. *Game Sound Technology and Player Interaction*. IGI Global, 2011.

—. *Designing Sound*. Cambridge: The MIT Press, 2010. Print.

—. *Obiwannabe*. n.d. 8 February 2014. <http://obiwannabe.co.uk/html/papers/proc-audio/node19.html#SECTION00061200000000000000>.

Hoffert, Paul. *Music for New Media*. Boston: Berklee Press, 2007.

Holman, Tomlinson. *Sound for Digital Video*. Burlington: Focal Press, 2005.

Howard, David M and Jamie A.S. Angus. *Acoustics and Psycoacoustics*. Oxford: Focal Press, 2009.

http://www.cnet.com/profile/sonyinsider/. “Deathly din: How Dead Space 3 sounds are made.” 2013 . Web. 09 June 2013. <http://www.cnet.com/profile/sonyinsider/>.

Huiberts, Sanders. “Captivation Sound:The Role of Audio for Immersion in Computer Games.” Dissertaion. 2010.

Husserl, Edmund. *The Phenomenology of Internal Time Consciousness*. Indiana University Press, 1964.

Izhaki, Roey. *Mixing Audio:concepts, Practices and Tools*. Burlington: Focal Press, 2008.

Jørgensen, Kristine. “Computer Game Audio and Player Action.” PhD Thesis. 2007. PDF.

Jørgensen, Kristine. *Game Studies*. n.d. 10 January 2014.

Jørgensen, Kristine. “Left in the Dark: Plaing Computer Games with the Sound Turned Off .” Collins, Karen. *From Pacman to Pop Music*. Aldershot: Ashgate Publishing, 2008.

Jørgensen, Kristine. “On transdiegetic sounds in computer games.” *Northern Lights Volume 5* 2005. Print.

Kaae, Jesper. “Theoretical Approaches to Composing Dynamic Music.” Collins, Karen. *From Pac-man to Pop Music*. Burlington: Ashgate, 2008. 78.

Kuehnl, Eric. *Granular Synthesis*. n.d. 25 January 2014.

Liljedahl, Mats. “Sound for Fantasy and Freedom.” Grimshaw, Mark. *Game Sound Technology and Player Interaction*. n.d.

Marks, Aaron and Jeannie Novak. *Game Audio Development*. Clifton Park: Delmar, 2009. Print.

Marks, Aaron. *Game Audio*. Clifton Park: Delmar, 2009. Print.

Moylan, William. *Understanding and Crafting the Mix*. Burlington: Focal Press, 2007.

Mullan, Eoin. “Physical Modellling for Sound Synthesis .” Grimshaw, Mark. *Game Sound Technology and Player Interaction*. Hershy: IGI Global, 2011. 340.

Murch, Walter. “Foreword.” Chion, Michel. *Audio Vision: Sound on Screen*. New York: Columbia University Press, 1994.

—. “Sound Design: The Dancing Shadow.” *Projections 4: Filmakers in Filmaking* 1995.

—. *Stretching Sound to Help the Mind See*. n.d. January 2014. <http://filmsound.org/murch/stretching.htm>.

Nacke, Lennart and Mark Grimshaw. *Game Sound Technology and Player Interaction*. Hershey PA: Information Science Reference, 2011.

Paul, Leonard J. “An introduction to Granular Synthesis in Video Games.” Collins, Karen. *From Pac-man to Pop Music*. Aldershot: Ashgate Publishing, 2008.

Price, Simon. *Granular Synthesis: How It Works & Ways To Use It*. December 2005. 25 January 2014. <http://www.soundonsound.com/sos/dec05/articles/granularworkshop.htm>.

Roads, Curtis. *The Computer Music Tutorial*. Cambridge, Massachusetts: The MIT Press, 1996.

Rona, Jeff. *The Reel World: Scoring for Pictures*. San Fransisco: Miller Freedman, 2000.

Roux-Girard, Guillaume. “Listening to Fear: A Study of Sound in Horror Computer Games.” Grimshaw, Mark. *Game Sound Technology and Player Interaction*. IGI Global, 2010.

Russ, Martin. *Sound Synthesis and Sampling*. Oxford: Focal Press, 2009.

Sonnenschein, David. *Sound Design: The Expressive Power of Music, Voice and Sound Effects in Cinema* . Studio City: Michael Wiese Productions , 2001. Print.

Stevens, Richard and Dave Raybould. *The Game Audio Tutorial: A Practical Guide to Sound and Music for Interactive Games*. Burlington: Focal Press, 2011. Print.

Thom, Randy. *Designing A Movie For Sound*. 1999. 26 January 2014.

Toole, Floyd E. *Sound Reproduction: Loudspeakers and Rooms*. Burlington: Focal Press, 2008.

Van Geelen, Tim. “Realizing Groundbreaking Interactive Music.” Collins, Karen. *From Pac-man to Pop Music*. Aldershot: Ashgate Publishing Limited, 2008.

Whittington, William. *Sound Design and Science Fiction*. Austin: University of Texas Press, 2007. Print.

Zizza, Keith. *Audio in Game Development*. n.d. PDF.

Ludography

Irrational Games. *Bioshock Infinite*. 2K Games, 2013. PC.

Bioware. *Mass Effect 3*. Electronic Arts, 2012. PS3.

Visceral Games. *Dead Space 3*. Electronic Arts, 2013. PS3

Eidos Montreal. *Deus Ex: Human Revolution*. Square Enix, 2011. PC

Volition Inc. *Saints Row 4*. Deep Silver, 2013. PC

Arkane Studios. *Dishonered.* Bethesda, 2012. PS3

Crytek. *Crysis 3*. Electronic Arts, 2013. PC

Valve Software. *Portal 2*. Valve Corporation, 2011. PC

4A Games, *Metro Last Light*. Deep Silver, 2013. PC

Starcraft 2 (2010),

Filmography

Alien (1979),

Blade Runner (1982)